Bias, Structure, and Injustice: A Reply to Haslanger

Robin Zheng
Yale-NUS College, robin.zheng@yale-nus.edu.sg

Bias, Structure, and Injustice: A Reply to Haslanger[1]
Robin Zheng

Abstract

　　Sally Haslanger has recently argued that philosophical focus on implicit bias is overly individualist, since social inequalities are best explained in terms of social structures rather than the actions and attitudes of individuals. I argue that questions of individual responsibility and implicit bias, properly understood, do constitute an important part of addressing structural injustice, and I propose an alternative conception of social structure according to which implicit biases are themselves best understood as a special type of structure.

Keywords: structural injustice; implicit bias; moral responsibility; social justice

I. Introduction

　　Moral philosophers, those who write on matters of morality and those concerned to conduct themselves morally, have taken great interest of late in the phenomenon of implicit bias, for many reasons. To give just two: implicit biases have been used to explain the persistence of social inequalities (e.g., the underrepresentation of women and minorities) long after discrimination has been outlawed, and to advocate changes in everyday practices that may be implicitly biased (e.g., grading, hiring, peer-review) (Saul 2013a). However, it has long been argued that oppression is not reducible to problems with individuals or individual psychologies. In other words, we might say, "Racism (sexism, etc.) ain't in the head." Rather, forms of oppression are complex systems of disadvantage encompassing

---

political, economic, social, and cultural institutions—in short, they are structural (see, e.g., Collins [1990] 2000; Frye 1983).

What is the connection between implicit bias and structural injustice? Sally Haslanger (2015) has made a crucial intervention by posing this question. Her answer, however, is largely negative: "If the best explanation of social stratification is structural, then implicit bias seems at best tangential to what is needed to achieve justice. Why the recent emphasis on implicit bias as a solution?" (2). Haslanger's critique raises challenges for the burgeoning moral literature on implicit bias (see, e.g., Brownstein and Saul 2016), which has predominantly been concerned with the question of individual responsibility for biases. Specifically, Haslanger charges that the individualism of the implicit bias literature makes it "inadequate to explain ongoing injustice" and "fails to call attention to what is morally at stake," and hence that "an adequate account of how implicit bias functions must situate it within a broader theory of social structures and structural injustice" (1).

In this paper, I take some steps toward carrying out that task by showing that questions of individual responsibility and implicit bias, properly understood, do constitute an important part of addressing structural injustice. In Section II, I present Haslanger's argument against focusing on implicit bias, which falls out of a deeper critique of individualist approaches to injustice. I first address the charge of individualism in Sections III and IV. I defend the need for a theory of individual responsibility for structural change, and I show how recent theories of responsibility as *accountability* provide a different way of focusing on individual action and attitude that is responsive to Haslanger's concerns. I further argue that collective action itself requires individualistic responsibility practices, using examples from contemporary labor and social movements. Finally, in Section V, I briefly consider implicit bias on its own terms, suggesting an alternative understanding on which biases are *themselves* a certain kind of social structure. In short, social transformation begins once we step out from the explanatory perspective of social theorist and into the practical perspective of organizer or concerned individual, and it is from this vantage point that individual responsibility and implicit bias loom large.

II. Haslanger's Challenge
Explanatory vs. Normative Individualism

Haslanger's critique, in a nutshell, is that if rectifying structural injustice cannot be achieved by modifying individual (racist, sexist, etc.) actions and attitudes, then *a fortiori* it cannot be achieved by modifying *implicit* (racist, sexist, etc.) attitudes. She identifies and rejects two potential roles implicit bias might play in theorizing injustice: an explanatory role identifying the causes of structural injustice, and a normative role identifying what is wrong with and who is responsible for

structural injustice. The problem with both, for Haslanger, is their *individualism*, their focus on individuals rather than social structures.

Haslanger's primary target is the presumed explanatory role of implicit bias and its use in "standard stories." Standard stories are explanatory narratives in which social outcomes are explained as the aggregative effects of individual agents' actions—for example, when racial and gender inequality are explained as the cumulative result of individuals' (implicit) racist and sexist decisions. But, as Haslanger demonstrates, racist and sexist attitudes are not necessary for unjust outcomes, for there are many that result from ordinary decent individuals acting in perfectly reasonable ways, given their situations: a husband and wife whose decision to make her the primary caregiver (because only she receives parental leave) results in unequal incomes, a teacher whose fair disciplining of a Black student leads him and his non-White friends (because they have experienced long patterns of racism) to disengage and perform badly in her classroom, and a worker who loses his job due to the city's cancelling (because they are strapped for cash) his bus route to work. In these cases, individuals might have sexist or racist attitudes, or they might not. But just giving an account of structural constraints in each case is sufficient to make intelligible why these individuals acted as they did and what resulted; no further investigation into whether they also harbor implicit biases is necessary.

My primary aim in this paper, by contrast, is to consider individualism and implicit bias from a normative perspective—that is, to defend a "normative focus" on individuals (Haslanger 2015, 10). Although Haslanger is not very clear on what she means by this, she seems to countenance at least two distinct ways in which individual action and attitude might serve as a focus in normative theorizing: first, as a guide to the best explanation of what is wrong with injustice,[2] and second, as a guide to who is responsible for injustice.[3] She rejects the first possibility for the following reasons:

> 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. It reinforces fictional conceptions of

---

[2] Early in the paper she asserts that the "*normative core* of what is wrong with racism/sexism lies not in the 'bad attitudes' of individuals but in the asymmetrical burdens and benefits and inegalitarian relationships that societies impose on such groups" (Haslanger 2015, 2).

[3] In the same section, she considers the view that "implicit bias may be part of the normative story, for insofar as we can change our implicit attitudes, we are plausibly responsible for them" (2).

autonomy and self-determination that prevents [*sic*] us from taking responsibility for our social milieu (social meanings, social relations, and social structures). (10)

The idea seems to be as follows. Responsibility for an action, traditionally understood, requires that an agent was able to choose differently than she did (or that she had control over her actions, did so voluntarily, and so on). By trying to hold individuals responsible for actions contributing to injustice, we thereby ascribe to them forms of autonomy and self-determination that they lack. In reality, their choices are highly constrained by social structures.[4] As Haslanger puts it: "Structures, and their component schemas and resources, can be responsible for injustice, without implicit bias or ill-will on the part of the participants in a milieu" (4).[5] The wrongness of injustice, then, consists primarily in bad social structures rather than bad individuals.

Moreover, since individuals lack the power to change social structures, Haslanger advocates shifting normative focus away from individuals to social groups, that is, shifting away from a "tendency to focus on the individual (or state?) as the moral (or political?) agent" to a "politics of claim-making by and on behalf of social groups" (10–11). She writes:

> Social change requires contestation, organization, and activism. Working to correct our own biases may be a minimum requirement on us. But we are each complicit in the perpetuation of unjust structures, practices, and institutions. Moral responsibility concerns not only what I can and should do, but also what we can and should do together. (12)

I wholeheartedly endorse Haslanger's call to collective contestation. However, I am doubtful that we can[6] or should avoid a normative focus on individuals, in the sense

---

[4] See Lavin (2008) for a broadly similar argument.

[5] Indeed, in earlier work Haslanger has defended a distinction between "agent oppression," an act of wrongful harm committed by an individual, and "structural oppression," that is, cases in which "there may not be an oppressor, in the sense of an agent responsible for the oppression [because] practices and institutions oppress" (2004, 103). Rather than think that no agent is responsible for injustice, my view is that we should think of *all* agents as responsible (but only in the accountability, rather than attributability, sense; see Section III).

[6] Perhaps this could be done in terms of collective responsibility; indeed, many philosophers, following Virginia Held (1970), have conceived of responsibility for injustice as a problem of *collective inaction*. On this view, individuals can be judged

of viewing them as the appropriate bearers of responsibility for structural transformation. In particular, I think that the two kinds of normative focus come apart. While questions of normative explanation can be undertaken from a broadly theoretical third-personal perspective (as Haslanger does), questions of responsibility are necessarily addressed from the first- and second-personal *practical* perspectives of motivating and responding to action, from which we cannot but think of ourselves and others as persons imbued with individual agency.

It is all very well to say that we need structural solutions rather than reformed individuals, but it is much less obvious what kind of collective action should be undertaken and how. Existing social structures are highly resilient and exert strong pressures on individuals, who often prefer to simply try surviving the status quo. As many organizers would attest, the biggest obstacle to collective action usually just *is* empowering people to think collectively rather than individually—to invest substantial time, effort, and hope in the face of dauntingly powerful structures of domination. Moreover, it is often those most disadvantaged by existing structures who possess the fewest resources to spare toward political organizing. Under these conditions, then, how can we hold people responsible for organizing collective action? What does it actually mean for us to take responsibility for our social milieu?

In the next two sections I shall argue that the best answers to these questions, *especially* when we accept Haslanger's call to collective action, retain a normative focus on individuals. Specifically, I shall argue that theories of individual moral responsibility for injustice are available and necessary (Section III), and that collective organizing itself must involve certain individualistic accountability practices (Section IV).

---

blameworthy for *failing* to form a hypothetical collective agent that they could have formed (Held 1970). Wary of the charge that collective responsibility is "barbarous" for punishing individuals for the actions of others, however, theorists of collective responsibility have worked to preserve the same conditions of autonomy and self-determination required for responsibility as traditionally understood. Thus with regard to collective inaction, they have limited ascriptions of blameworthiness to cases in which it would be clear to a reasonable person that a collective agent could and should be formed (Isaacs 2011) or to instances when there is sufficient time and knowledge of how to coordinate actions such that there is "practical plausibility to the counterfactual claim that the group could have done otherwise" (May 1990, 269–277). I am doubtful, however, that these clarity and plausibility conditions are often or ever met (cf. Doan 2016).

III. Accountability for Structural Injustice
From Attributability to Accountability

  The problem, I think, is this: extant theories of responsibility for implicit bias—Haslanger's target—have largely been theories of responsibility as *attributability*, whereas the most promising theories of responsibility for injustice are theories of *accountability*. While the purpose of attributability is to assess people's quality as moral agents, the purpose of accountability is to distribute duties and burdens across a moral community. The problems Haslanger identifies with individual responsibility are features of attributability but not accountability. While the former may be orthogonal to structural injustice, the latter is important and well-suited for rectifying injustice.

  In earlier work I advanced an interpretation of the attributability vs. accountability distinction as originating from two separate philosophical problems (Zheng 2016). Attributability has its roots in metaphysics and action theory, and arises from the problem of action: *What makes my actions count as genuine exercises of my own agency (and not merely events happening to or within me)*? The primary function of this concept is delineating which actions count as legitimate exercises of agency and hence as grounds for assessing moral agents as good or bad exemplars of their kind; accordingly, the conditions for ascribing attributability are typically psychological or metaphysical in nature.

  The implicit bias literature has focused on just this question of whether implicitly biased actions are properly attributable to agents as exercises of their moral agency. Jules Holroyd, for instance, argues that we have the types of control and awareness required for responsibility—namely, indirect long-range control over actions that can mitigate bias over time (Holroyd 2012), and occurrent observational awareness that one's responses are biased (Holroyd 2015). By contrast, Neil Levy (2014a, 2014b) cautions against ascribing responsibility for implicit biases because they are only "patchy endorsements" and cannot be taken to reflect an agent's evaluative stance. I have argued similarly that biases are not attributable when agents would not reflectively endorse them and have otherwise made reasonable efforts to mitigate them (Zheng 2016). But Michael Brownstein contends that we are responsible for implicit biases because they express our *cares*, that is, what really matters to us, even if we do not endorse them (Brownstein 2015). Each of these views proposes some candidate condition—control, awareness, evaluative endorsement, and expressing care—under which an implicitly biased action may be attributed to a person as reflective of her agency and hence warrant blame, punishment, or reactive attitudes.

  However, a different concept of responsibility—responsibility as "accountability"—is more important for cases of implicit bias. Accountability is rooted not in metaphysics but in moral and political philosophy, where the

foundational question is, *How should we assign duties and burdens across the moral community in line with principles of justice and fairness?*[7] When a person fails to carry out some duty, the concept of accountability functions in the first instance to ensure that the moral community takes up the costs of dealing with harm done, by distributing those costs to appropriate parties. It is not *yet*, and not necessarily at all, to determine whether she has acted badly or well or what kind of agent she is. Of course, if the reason an agent has failed to carry out her responsibility is that she acted badly, then she deserves blame and punishment as well—but whether she does is a *further* question that we can, but need not, ask. Seeking to assign blame *in addition to* assigning burdens of repair will trigger considerations of attributability at this later stage, which is why the two concepts are often conflated. In other words, a person may be ascribed *accountability* for some action even if she is not *attributively* responsible for it, or if we do not care to further investigate whether she is also attributively responsible.

Theories of accountability have primarily been developed by political philosophers. Iris Marion Young (2011), for example, distinguishes between what she calls the "liability" and "social connections" models of responsibility. The former is a conception of attributability, on which an agent is responsible for an unjust outcome if she brought it about knowingly and voluntarily. Young, however, advocates that we adopt the latter model, a conception of accountability on which everyone who causally participates in an unjust outcome is responsible for bearing some of the burdens of collectively reforming the structural processes that led to it. Robert Goodin (1987) has defended a theory of "task-responsibility," according to which people are assigned different *ex ante* tasks, and responsibility is subsequently apportioned in accordance with a person's role, rather than evil intentions or character-revealing negligence; this distinguishes it from what Goodin calls "blame-responsibility"—another conception of attributability.

In light of the above, it is clearly theories of *attributability* rather than accountability that are vulnerable to Haslanger's critique. Whether an agent acted with autonomy and self-determination is relevant to determining if an action is properly *attributable* to her, because actions lacking those features do not count as genuine exercises of moral agency. They may not have been chosen under the agent's control, with her reflective endorsement, as an expression of her cares—or however we want to cash out "autonomy" and "self-determination" metaphysically

---

[7] Note that this refers to a *normatively* acceptable, i.e., just and fair, rather than *actual* distribution of duties and burdens. Injustice may consist precisely in the way that duties are inequitably distributed, e.g., how caretaking responsibilities fall heavily on women rather than men. I am grateful to an anonymous reviewer for pushing me to clarify this point.

and psychologically; it would be unfair to blame, punish, or resent her in these cases even if the action is morally faulty. But autonomy and self-determination are not required for assigning burdens to an agent for the purpose of remedying some harmful outcome. Since those costs must be distributed somehow or other, an agent may be fairly assigned these burdens merely on the basis of causal proximity, or in virtue of her role, even if she was in no way at fault. Thus, just as we can hold an agent accountable for her implicitly biased action without thereby attributing it to her, we can hold an agent accountable for structural injustice without attributing any fault to her.[8]

Why Responsibility?

At this point it might be asked, why do we need any concept of responsibility at all, be it attributability or accountability? Why not dispense with these "fictional" conceptions of autonomy and self-determination altogether, given that it is often structures rather than individuals that oppress?

I would like to offer three reasons why individual responsibility remains important for theorizing structural injustice. First, theories of responsibility do important normative work in *justifying* and *enforcing* the demand to work toward structural transformation. For we have accepted that unjust outcomes result even when individuals act in reasonable ways that would otherwise be morally unproblematic, that is, if not embedded within overarching unjust social structural systems. If that is so, then why should any *particular* agent be expected to change their behavior,[9] so long as they do not commit blatant wrongdoing? As Young (2011,

---

[8] Indeed, both of my arguments for ascribing accountability rather than attributability for implicit bias transfer directly to structural injustice. Just as it is difficult to establish the operation of implicit bias in any given *particular* case (though it can be determined at the aggregate level), so too does the nature of structural injustice mean that it is almost impossible to identify precisely how a specific agent's actions directly caused a specific harm to others (though the harms of the overall structural process can be identified) (I. M. Young 2011, 96). And just as blaming people for implicit bias often provokes denial and resistance, structural injustice often produces what Young (2011, 117) calls a "round-robin 'blame game'" in which parties (justifiably) point to others' actions over which they had no control, rather than their own, as the cause of some unjust outcome.

[9] This problem is vividly summed up in a popular three-panel cartoon depicting a man speaking to a large audience from a podium. In the first panel, all hands are up and the crowd is all smiles, as he asks: "Who wants change?" In the second, all hands have come down and the smiles have vanished, as he asks: "Who wants *to*

165) notes, people often think the task of structural transformation is "not their job," but rather someone else's, for example, the state's. Some justification is required for demanding of individuals that *they* undertake burdens for the sake of rectifying injustice.

Ascribing responsibility to a person provides justificatory grounds for such a demand. When we ascribe responsibility to someone, we make a normative claim on them to the effect that *they* are expected to act in some way or take up some kind of burden, such that they are appropriately subject to critical moral responses by others. For attributability, the grounds for this claim lie in the fact that persons are moral agents and are expected to exercise their agency in accordance with moral norms. When they exercise their agency badly, they deserve on that basis to be blamed, punished, or subject to other reactive attitudes. For accountability, both justification and response will be different (as I explain in the next section). Of course, actual agents may take up the work of structural transformation out of all sorts of motives: compassion, benevolence, religiosity, rebelliousness, or ambition, to name a few. But bearing responsibility is a *normatively* powerful reason for acting.

Second, some notion of responsibility—with its attendant notions of autonomy, self-determination, and the like—is necessary for action. Haslanger notes that standard stories "may be irresistible for humans . . . because they focus on the autonomy of persons and enable us to locate and judge moral responsibility" (Haslanger 2015, 9). I would say, rather, than they are *indispensable*, because it is always from the individual first-person perspective that each of us acts. In acting, we cannot help but experience ourselves as having the choice to act or not act, in one way or the other. Conversely, experiencing this choice *as a choice* is a necessary step for undertaking any action at all, but especially action directed against what can feel like overwhelmingly oppressive social structures. As Haslanger notes, structure and agency are interdependent: while structures constrain individuals' actions, they only exist so long as they are enacted and reenacted by individual agents. The possibility of social change, then, lies in the agency of individuals within a structure whose behavior maintains it—whose behavior, therefore, might alter it. And it is usually other moral agents who alert us to the existence of this agency and shape our experience of it.

This means, finally, that theories of moral responsibility perform an important action-guiding function. They direct us, from the second-personal perspective, toward morally appropriate responses to others' actions. Thus, normative individualism might not derive from the "impoverished" view that

---

change?" In the final panel, he is left facing an empty room after he asks: "Who wants to *lead* the change?"

Haslanger rejects, that is, the individualist idea that "our only (or best) option for changing social reality is to take responsibility for our own thought and perception and encouraging others to do the same" (2015, 12). Rather, theories of moral responsibility provide guidance in cases where others fail to take up their share of the collective burden or make mistakes in doing so—both of which are unavoidable on the long hard road to justice. While the work of structural change is collective, it is always particular individuals, their actions and attitudes, that we confront in the classroom, in the meeting hall, and on the streets. A theory of individual moral responsibility for structural injustice thus takes seriously the interpersonal relationships between persons that are key to the actual day-to-day work of contestation, organization, and activism. And different theories of individual moral responsibility as expressed in organizational practice, for example, whether attributability or accountability is being ascribed, can markedly impact how well the organization is able to engage its own members and to build coalitional solidarity with others. I illustrate these claims in the next section.

IV. Practicing Accountability in Collective Organizing

I have argued that we need a theory of individual moral responsibility for structural injustice, and that accountability is ascribable even when attributability is not, such as in contexts of injustice. But how should we actually hold agents accountable for injustice? While political philosophers have explicitly eschewed practices of blame associated with attributability, they typically do not offer any account of distinctive backward-looking critical moral responses that are central to any theory of moral responsibility.[10] In this section I describe two examples, organizing conversations and call-outs, that address problems central to the kind of collective organizing and activism that Haslanger advocates: the problem of mobilizing action, and the problem of maintaining solidarity.

The Burdens of Rectifying Injustice

First, however, I need to say something about how a theory of accountability addresses the problems of justification and response I outlined in the last section, since it will not depend on judging that a person exercised her moral agency poorly. In contexts of structural injustice, such ascriptions of attributability are problematic

---

[10] David Miller (2001, 454, 469) mentions "pressure of various kinds" and "sanctions," but states only that they will vary from case to case. Notably, even Christopher Kutz's (2000) careful treatment of the various relationship-dependent backward-looking responses merited by individuals complicit in collective harms changes tune when he reaches the problem of unorganized collectivities: there he redirects focus to cultivating forward-looking motivations.

because they mislocate what needs to be changed: the range of options available, not the agency of the individuals involved. Moreover, it is a highly fraught question just how much a person is morally obligated to do towards rectifying injustice, and we are often not in a position to know whether a person's failure to contribute on some particular occasion is due to a blameworthy exercise of agency or to other pressing structural constraints,[11] or whether she is contributing in some other way.

By contrast, normative claims on an agent based on accountability rest on very different justificatory grounds. To be sure, accountability does presuppose basic moral agency. But what serves as the real basis for normative claims here are the duties assigned to a person in virtue of her place within a moral community—something that, far from being inimical to a structural picture of injustice, fits right in. Indeed, we might combine Young's and Goodin's respective proposals for what is required to ascribe accountability—causal participation in structural processes, and role occupation—to say that individuals are accountable for working toward structural transformation in virtue of their *social roles*, that is, precisely *because* of their being embedded in certain positions within social structures.[12] In other words,

---

[11] Martha Nussbaum raises two objections against Iris Marion Young (2011), in favor of holding agents attributively responsible for not contributing. She argues first that "if A has responsibility R for social ill S, and she fails to take it up, then, when the relevant time passes, she is guilty of not having shouldered her responsibility," and second, that "if it is a general moral truth that citizens ought to monitor the institutions in which they live and be vigilant lest structural injustice occur within them, then I think it follows that they are culpably negligent if they do not shoulder that burden" (I. M. Young 2011, xxi). I believe, however, that this objection misconstrues the nature of injustice on Haslanger's and Young's structural accounts. Once we understand that injustice is structural, it becomes unrealistic to believe people ever *could* act on all the social ills for which they can be held in some sense responsible, or that they *could* adequately monitor the institutions they inhabit for possible injustice. Our institutions are *already* rife with injustice, through and through—after all, structural injustice by its very nature is not limited to particularly egregious cases of wrongdoing but comprises the very background conditions against which these cases occur. Moreover, there are so many dimensions of structural injustice that it is unclear how we can be considered to have fully and successfully "shouldered" our responsibility, so long as injustice persists. In other words, there is no "relevant time" at which our responsibilities suddenly get activated; we are at all times already responsible for doing what we can to change unjust background structural conditions.

[12] I do not have the space here to defend this particular conception of accountability as grounded in social roles. But in traditional social theory, social roles have a long

the normative claim is justified by the fact that each of us, *no matter how well we exercise our agency*, still perpetuates unjust outcomes through our otherwise unproblematic ordinary behaviors; hence we must share some of the burdens of rectifying that injustice by working to transform those behaviors.

The logic of accountability practices is thus importantly different from the logic of attributability practices. Blame, punishment, and the reactive attitudes consist of some negative appraisal of faulty agency, along with some additional unpleasant treatment or attitude that functions as negative sanction,[13] except where that sanction is unwarranted because the negative appraisal is undermined, that is, in cases where the agent has an excuse because the action did not count as a genuine exercise of moral agency. Practices of accountability, by contrast, do not require this kind of appraisal in order to be justified. Instead, I suggest, in the context of rectifying injustice they have the logic of *reminders*—reminders of the fact that because one participates unavoidably in unjust structures,[14] one therefore bears a burden to work toward structural transformation. Reminders, I claim, are critical moral responses because their basic function is to redirect attention, to call upon an agent to stop and reflect on what else she ought to be doing.[15] Because this fact is not grounded in some poor exercise of agency—and, moreover, because the work of structural transformation is never over, so long as injustice persists—no appraisal of an individual's actions is necessary for such a reminder to be warranted. Reminders are appropriate, indeed, even when a person is already doing everything as she should, for example, when event reminders are sent to everyone on a mailing list, whether or not they are likely to forget. In other words, justification for accountability practices does not depend on an answer to the question "Have you acted badly?" but merely on the fact that it is more or less[16] always appropriate for us to be reminded of our burden. Let me be clear, however, that I do not mean to say that attributability practices of blame and the reactive attitudes are *never* warranted or efficacious; my point here is merely that accountability licenses

---

history of being understood as the interface between structure and agency (Dahrendorf 1968; Parsons [1951] 1991).

[13] There is considerable disagreement as to precisely what this additional negative sanction is. See Coates and Tognazzini (2013).

[14] Again, I do not defend a particular conception of accountability here, but different conceptions will identify different facts *in virtue of which* one bears this burden: causal participation in structural processes, role occupation, etc.

[15] See Springer (2013) for an account of critical moral responses as functioning to communicate moral concern about some problem.

[16] By this I mean that it is always *pro tanto* morally justified; it may, for instance, be inconsiderate or insensitive during moments of grief, healing, and so on.

different critical moral responses that may often be justified or effective in cases where blame is not.

It might be objected here that it is unfair for the burdens of rectifying injustice to fall also on its victims, since, as I mentioned in Section II, they have the least resources to spare toward collective action. But this objection lapses back into the framework of attributability. As Young (2011, 146) points out, while it would be "perverse" to judge victims of injustice blameworthy for their condition (as on an attributability model), they still share responsibility (as accountability) for these conditions in the sense that they can be reasonably expected to take steps to collectively organize against them, if only because they know the most about their own oppression. Moreover, the objection rests on an overly narrow assumption of the ways in which the burden of rectifying structural transformation can be discharged. Some of the most visible ways of contributing to structural transformation—protesting in the streets, lobbying city councils and state legislatures, or overhauling discriminatory institutional policy—certainly are burdensome, and it is thus particularly important for those in positions of privilege and power to support this work. But contributions need not all be like this. For structural transformation requires not only changes in material conditions and legal institutions, but also in culture and ideology (Haslanger 2017). There are thus many different ways, formal and informal, in which people with varying amounts of time and resources may work to discharge their transformative duties: not only in workplaces and unions, but also in their own churches, neighborhoods, households, and so on. Each of these relationships[17] and contexts presents opportunities for people to facilitate (local) structural transformation, in ways that need not be materially burdensome.

For example, consider what Patricia Hill Collins (2000) calls the "motherwork" performed by Black women, a contribution performed outside of a formal institutionalized role. Collins points out that Black women fight oppression when—as an alternative to riskier, more visible forms of activism that require greater time and resources—they "resist passing on to their children externally

---

[17] A striking fact about the astonishingly rapid shift in attitudes toward LGBT people is that those who report to know many LGBT people are more than twice as likely to support gay marriage as those who do not, and those who report having an LGBT close friend or family member are 1.75 times more likely; the effect of personal acquaintance is significant even after controlling for demographic factors (Jones, Cox, and Navarro-Rivera 2014). In other words, merely *being* in a relationship can sometimes constitute meaningful political work. And although such attitude change is no substitute for structural reform, it is important for initiating, implementing, and preserving structural change against backlash (see Madva 2016).

defined images of Black women as mules, mammies, matriarchs, and jezebels" and choose instead to "use their families as effective Black female spheres of influence to foster their children's self-valuation and self-reliance" (210). Because of the way that "political consciousness can emerge within everyday lived experience" (209), all individuals have ample opportunity to contribute, however incrementally or imperceptibly, to the collective project of structural transformation. The particular *ways* in which they do so will vary widely depending on their position within a social structure, and in particular on their social roles, because different social roles confer different powers and enable different relationships for individuals who occupy them.[18] For some, the effort of merely surviving with one's self intact represents a form of resistance against hostile structures built to crush them; as Audre Lorde famously wrote: "Caring for myself is not self-indulgence, it is self-preservation, and that is an act of political warfare" ([1988] 2017, 130). Thus the nature of each individual's contribution will be different. But we all share equally, together, the collective burden of structural transformation—and this fact justifies us in deploying critical moral practices of accountability to remind each other of that burden.

The Problem of Mobilizing Action: Organizing Conversations

The two practices of accountability I explore here address problems inherent to any kind of collective organizing. My first example, the organizing conversation, is a response directed toward omissions or failures to act in certain ways; it promotes *taking* responsibility (as accountability) for structural injustice. My second example, the call out, is a way of *holding* others responsible for actions that reproduce oppressive dynamics amongst allied groups. In both cases, a theory of individual responsibility provides an underlying rationale for guiding second-personal responses to others' inaction and mistakes.

Perhaps the largest barrier to collective organizing (to which I alluded in Section I) is the problem of mobilizing people to act against apathy, fear, limited time and resources, competing obligations, and the many other reasons that people fail to participate in collective political action. Such failures to act, I claim, do warrant critical moral response—but not blame. One example of the kind of response I have in mind can be found in the contemporary labor movement, which has long been an exemplar of successful collective organizing. Many unions train their members to go door-to-door and engage in face-to-face conversations with other members, to persuade them to be more active. A paradigm "organizing conversation" unfolds as follows. The organizer, instructed to spend 70 percent of

---

[18] See Iris Marion Young (2011) for an account of the parameters (power, privilege, interest, and collective ability) that alter how differently positioned people can contribute to structural transformation.

the conversation listening and 30 percent talking, begins by asking the worker to describe who she is and what her work experience is like (Weingarten, Cortese, and Johnson 2009). The next step is to "agitate and educate" by asking questions such as, "Who decides? Why are things the way they are, who has the power to determine working conditions and policies, what role do unit members have in determining work and professional conditions?" (49). This is followed by "the union vision," in which the worker is asked to imagine an ideal workplace through questions such as, "What would it take/what would need to happen for you to do a better job/better serve students/provide better patient care" (49). Both of these steps require the worker to locate specific individuals and herself within the social structure of the workplace, along with a range of possible actions that could be performed from those respective positions. The conversation thus serves to highlight the individual agency that sustains—and can therefore alter—the structure.

Next comes an inoculation against common fears and objections. If an individual describes herself as unable to do anything about a problem, for instance, the organizer can ask what would happen if her entire unit did something together, such as sign a joint letter or show up together in their supervisor's office. Finally, the organizer is in a position to make an "ask.". What is crucial about making asks is *meeting people where they are*, no matter where that is, and pulling them just a bit further—to schedule a follow-up conversation, provide information, join a mailing list, wear a button, attend a meeting, or recruit another member—because "that's how 3s become 2s and 1s" (49).

Whilst the organizing conversation is designed specifically for union-building, it seems clear to me that it is also a practice of holding people accountable for *taking responsibility* for injustice. There is a critical dimension to the contours of the conversation, because it functions as a reminder of the agency that individuals can exercise in service of structural transformation—where that agency can be exercised differently by different individuals and might consist in something quite small, such as wearing a button. The organizing conversation thus encourages two senses of "taking responsibility" identified by Claudia Card: a person's "agreeing to answer or account for something, or finding that one should be answerable" along with "committing [themselves] to stand behind something, to back it, support it, make it good" (Card 2010, 28). But its internal logic is *not* justified by any kind of agent appraisal; organizing conversations are directed at an entire membership, not only inactive members. The goal is not to negatively sanction anyone, but to acknowledge the constraints individuals face while reminding them of what they can do and empowering them to do what they can. It is therefore important, as

emphasized by unions using the "organizing" rather than "service" model,[19] to engage in regular and consistent organizing, outside of particular campaigns or in reaction to particular abuses of power (Banks and Metzgar 1989). One can imagine the organizing conversation being adapted for a variety of problems, or, more likely, one can recognize trace elements of it in all sorts of ordinary conversations that already transpire between concerned individuals. Although the organizing conversation is not the only way to mobilize action, it is the most reliable and effective technique in the organizer's handbook.

The Problem of Maintaining Solidarity: Call-outs
        A second significant barrier to collective action is the difficulty of preserving solidarity amongst diverse groups who relate to one another across varying dimensions of oppression. The White working class is oppressed, but they may oppress their fellow Black workers; Black men are oppressed with respect to race, but they may also oppress Black women, and so on—this is a familiar point (e.g., hooks 1984). The easiest way to prevent collective action is for such groups to be divided and conquered.[20] Contemporary social movements organizing around identities such as race, gender, sexual orientation, disability, etc. represent the cutting edge in thought and praxis concerning the day-to-day practicalities of intergroup solidarity. One example of such a practice is the "call-out." To "call out" a person is to point out to them how something she has said is racist, sexist, heterosexist, ableist, or otherwise oppressive. Call-outs represents an important way in which "allies," that is, people who do not themselves belong to but support the liberation of some oppressed group, can work to relieve some of the onus[21] that usually falls on members of those groups.

_____

[19] "Service model" unions focus on providing expert services on behalf of their members, e.g., negotiating labor contracts and providing legal representation. "Organizing model" unions, by contrast, emphasize high involvement and activity on the part of rank-and-file members, not just the leadership.
[20] While exogenous factors such as violent repression can certainly also crush a movement, I focus on the dangers of internal divisions because they are inherent to all attempts at mass organizing, and because they are in principle preventable by those participating in such efforts. For accounts of how racial and gender divisions have weakened the labor movement see, e.g., Fletcher and Gapasin (2008) and Cobble (2007). Cobble (2007) and Clawson (2003) argue that the revival of the US labor movement will require prioritizing the concerns of diverse constituencies concentrated in the new identity-based social movements of the 1960s and 1970s.
[21] This is what Nora Berenstain (2016, 570) calls "epistemic exploitation," that is, "the unpaid and often unacknowledged work of providing information, resources,

What I want to point out here is that critical moral responses like call-outs are important—*especially* in the context of collective organizing—because there is a normative and practical problem here that structural explanations cannot automatically resolve. The specific individuals involved face the real, live question of how to feel and relate to one another after what has just been said or done. How *should* one feel and respond after oppressive behavior? Does it matter if it was done unwittingly and with good intentions? What are the psychic and other costs of educating or carrying on with that person—for oneself and for the group?

These questions are part of an acrimonious and ongoing debate in (online) activist circles around what has been deemed "call-out culture." Critics of call-out culture charge that the practice produces a "toxic" climate of vicious attacks, banishment and ostracization, self-censorship, fear, and ultimately the disintegration of the movement (Ahmad 2015; Lee 2017; Trần 2016). Defenders of call-outs, however, point out that these criticisms have been largely aimed at women of color, and in return raise charges of the "tone argument" or "tone policing," that is, the "act of disregarding the substance of someone's argument by focusing on the way it was conveyed" (C. Young 2014). The argument here is that objections to being called out often amount to (privileged) people not being able to tolerate being held accountable for their mistakes, and unfairly asking (marginalized) people to suppress their legitimate feelings of anger and resentment (Cooper 2014).

My aim is not to try and resolve this debate, but to emphasize again that the *practical* problem at hand, which has as a matter of fact generated significant divisions within the feminist activist community, is how particular *individuals* ought to engage with each other when recurrent oppressive dynamics in their midst threaten to compromise their ability to work collectively. I suggest that an important open question is whether we should understand call-outs as practices of attributability or accountability (where this will likely vary across situation, types of relationships, and so on). In other words, should we understand call-outs as functioning as a kind of negative sanction that a person deserves in response to oppressive behavior, or as an appraisal of an agent's oppressive attitudes (e.g., "calling out a person's ableism")? Or should we understand them as assigning burdens of repair, that is, for the person called out to apologize for the psychological harm she has caused and to take up the work of educating herself and others? My own view is that the latter is preferable.[22] In any case, theories of attributability and

---

and evidence of oppression to privileged people who demand it—and who benefit from those very oppressive systems about whose existence they demand to be educated."

[22] It seems to me that many of the damaging effects of call-out culture are traceable to people's tendency to use call-outs to ascribe attributability. Katherine Cross

accountability, by delineating the conditions required for different types of responsibility, can help us to navigate this complex moral terrain as it manifests across varied social contexts, such that we respond to each other in morally appropriate (and practically efficacious) ways as we organize ourselves for collective action.

I take it, then, that at least one of the aims behind the moral philosophical project of theorizing individual responsibility for implicit bias is to aid in situations like these. As I stated in Section III, a normative focus on individuals need not be motivated by the idea that we are each only responsible for our own psychologies. Rather, it can reflect a concern for the ways that we as social beings necessarily relate to one another *qua* particular individuals. Because my interlocutor is always a *person*, not a structure, I cannot dispense with perceiving actions and attitudes as manifested in particular individuals with whom I negotiate an ongoing personal relationship. And it is always a specific individual whom I engage in an organizing conversation, of whom I make a specific "ask" and encourage to undertake a specific action. A normative focus on individuals remains practically necessary insofar as an important part of the work of social change consists precisely in efforts to make normative claims on others that can elicit transformative agency. Even the most well-oiled political machine is subject to the vagaries of interactions between specific individuals both at the "top" within its leadership core as well as at the "bottom" peripheries wherein lies the hard work of recruitment and turnout.[23] Thus, the kind of interpersonal work expressed paradigmatically in practices of responsibility forms a necessary part of the task of *actually carrying out* collective action.

(2014) describes her fear of "being cast suddenly as one of the 'bad guys' for being insufficiently radical, too nuanced or too forgiving, or for simply writing something whose offensive dimensions would be unknown to me at the time." Asam Ahmad (2015), similarly, writes that the problem with call-out cultures is that "one action becomes a reason to pass judgment on someone's entire being . . . and to banish and dispose of individuals." As a practice of accountability, by contrast, call-outs might instead be executed in accordance with the popular ground rule: "Criticize the action, not the person," or with explicit refusal to attribute to persons the mistakes and imperfections that are inevitable in the process of "radical unlearning" (Lee 2017; Trần 2016). On the other hand, call-outs ascribing attributability may be important in response to certain kinds of behavior, like sexual harassment, where the conditions for counting such actions as exercises of genuine moral agency are more likely to be satisfied.

[23] For a fascinating study of such interactions in the Civil Rights movement, see Barbara Ransby's (2003) magnificent biography of Ella Baker.

Put differently, I have argued here that practices of individual responsibility are necessary to "get the ball rolling," as it were, and to "keep the ball rolling," because they are normatively grounded ways of getting individual agents to do things, and we cannot get a handle on a "structure" apart from getting at some individual or group of individuals. Before I conclude this section, however, let me clarify that I am *not* claiming that structural injustice would be rectified if only everyone was adequately held accountable for collective action. I am not even under the illusion that structural injustice can be rectified through labor unions, social movements, and collective organizing alone. If history is any guide, radical social transformation requires a confluence of factors, a perfect alignment of stars—many of which cannot be anticipated or forced. Yet significant social change can also occur gradually in the absence of great ruptures (e.g., growing acceptance of LGBT and transgender rights). More importantly, the steady day-to-day organizing work performed by labor unions, activists, and social movements, is *one* critical factor that must obtain. One cannot force a revolution, but one can try to be ready for one.

V. Implicit Bias, Habitus, and Action

Having defended a normative focus on individuals, along with the importance of responsibility practices in collective organizing, let me return now to the subject of implicit bias. The point I want to make is that implicit biases are not ordinary attitudes, and addressing biases can be part of structural transformation if we understand them to be a particular type of social structure.

Toward a Bourdieusian Account of Implicit Bias

Haslanger writes that the problem of structure versus agency is "arguably . . . *the* theoretical issue occupying social theory for the last three decades" (Haslanger 2015, 13). However, Haslanger's own conception of social structures as networks of practices, schemas, and resources (in the tradition of Anthony Giddens and William Sewell) is only one out of several social-theoretic conceptions of social structure. Here I consider how an alternative conception might accommodate the phenomenon of implicit bias, generating different implications for the explanatory and normative relevance of bias.

Pierre Bourdieu has famously laid claim to transcending the problem of structure versus agency altogether (Bourdieu 1990; Bourdieu and Wacquant 1992). While I cannot develop a fully detailed Bourdieusian theory of implicit bias here or do justice to the entirety of Bourdieu's field theory, let me sketch a sort of "proof of concept" grounded in his notions of *field* and *habitus*. Bourdieu conceives of social structures as *fields*, that is, as configurations of relationships between social positions. Agents can be mapped to locations in a field according to how much and what sorts of capital (i.e., social, material, and cultural resources) they possess; an

agent's trajectory within the field predicts her behavior (Bourdieu and Wacquant 1992, 94–100). As for *habitus*, Bourdieu writes:

> [Habitus] ensures the active presence of past experiences, which, deposited in each organism in the form of schemes of perception, thought and action, tend to guarantee the "correctness" of practices and their constancy over time, more reliably than all formal rules and explicit norms. (1990, 54)

Even on a first pass, this characterization of habitus is remarkably congruent with Greenwald and Banaji's (1995) early definition of implicit attitudes as the following:

> Implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects. (8)

Habitus both shapes and is shaped by the field; while habitus grows out of time within a field, fields persists only insofar as people remain invested in ways of acting that preserve them. There is thus a certain kind of mutually reinforcing *fit* that obtains between habitus and field and serves to maintains them both:

> Social reality exists, so to speak, twice, in things and in minds, in fields and in habitus, outside and inside of agents. And when habitus encounters a social world of which it is the product, it is like a "fish in water": it does not feel the weight of the water, and it takes the world about itself for granted. . . . It is because this world has produced me, because it has produced the categories of thought that I apply to it, that it appears to me as self-evident. (Bourdieu and Wacquant 1992, 127)

This description of implicit categories as introspectively invisible and taken for granted, yet efficacious in guaranteeing conformity to existing social realities, is again an impressively apt description of implicit social cognition. Moreover, Bourdieu's understanding of habitus as multifaceted, as "durably inscribed in the body and in belief" (1990, 58) through "schemes of perception, thought and action" (54), is well in line with recent philosophical theories of implicit associations as involving more than pure information processing. Tamar Gendler (2011), for instance, has argued that implicit biases are a species of alief, that is, "automatized representational-affective-behavioral triads" (41); Alex Madva and Michael Brownstein (2016) have also argued that implicit biases should be understood as clusters of semantic-affective associations straddling the divide between cognition and affect.

A crucial point for understanding Bourdieusian habitus is that habitus is *not* just another form of agency contrasted against structure (Lizardo 2004). This is important, because habitus (and hence implicit biases, if we conceive of them as an element of habitus) is not something attributable to a person as manifesting her agency and on the basis of which we can assess her quality as an agent. Or at least, it need not be construed as such, though that is one possible view.[24] For Bourdieu's is a structural approach on which the distinction between external forces acting on an agent versus an agent's own exercise of agency has no real meaning.[25] Yet the very problem of action that gives rise to questions of attributability depends on such a distinction to be meaningful.[26] Moreover, Bourdieu is characteristically uninterested in questions of individuals' blameworthiness, writing that "those who nowadays set themselves up as judges and distribute praise and blame . . . would be better occupied in trying to understand what it was that prevented the most lucid and best intentioned of those they condemn from understanding things which are now self-evident" (1977, 5).

Omar Lizardo (2004) thus interprets the habitus to be a type of structure[27] in the head that corresponds to structures outside the head (i.e., fields). He writes:

---

[24] I am grateful to an anonymous reviewer for pushing me to address this potential objection.

[25] He writes: "Overriding the spurious opposition of forces inscribed in an earlier state of the system outside the body, and the internal forces arising instantaneously as springing from free will, the internal dispositions—the internalization of externality—enable the external forces to exert themselves, but in accordance with the specific logic of the organisms in which they are incorporated" (Bourdieu 1990, 55).

[26] In his influential paper "Identification and Externality," Harry Frankfurt writes: "We think it correct to attribute to a person, in the strict sense, only some of the events in the history of his body. The others—those with respect to which he is passive—have their moving principles outside of him, and we do not identify him with these events. Certain events in the history of a person's mind, likewise, have their moving principles outside of him. He is passive with respect to them, and they are likewise not to be attributed to him" (1988, 61).

[27] This sort of view also seems to be available on Haslanger's preferred conception of social structure. For Haslanger, schemas are "intersubjective patterns of perception, thought, and behavior" that, along with resources, make up the practices constituting Haslangerian social structures (2012, 415); following Sewell (1992), she thus thinks of them as having only a "virtual" existence, i.e., as abstract objects. Habitus, by contrast, exists ontologically as part of an individual's psychology; what I am proposing here is that we understand implicit biases as

"*The habitus is itself an objective structure* albeit one located at a different ontological level and subject to different laws of functioning than the more traditional 'structure' represented by the field" (Lizardo 2004, 394, emphasis mine). This is what it means for social reality to exist "twice." To try and determine whether or not a person's habitus is attributable to her as a manifestation of autonomous or self-determining agency, then, is to lose sight of the fact that it is not that type of thing at all; it *is* social structure, but simply micro-level rather than macro-level.

This, I think, is a better way of thinking about implicit bias that does justice to why philosophers have been so captured by it. It is a crucial part of the explanation for social injustice that social structures get "deposited in persons in the form of lasting dispositions, or trained capacities and structured propensities to think, feel and act in determinant ways" (Wacquant 2005, 315). Hence implicit biases should not be assimilated into the category of individual actions and attitudes that Haslanger finds to be "beside the point" (2015, 8, 10). If I am right, they are no ordinary "attitudes" at all but precisely the thing that "fits" us to social structures, because they are themselves a species of social structure—one of the many interlocking and mutually reinforcing structures that constitute structural injustice.

Standard Stories Redux

Bourdieu locates social change in moments of crisis, where habitus and field fall out of alignment, such that previously invisible tenets encoded in the habitus are surfaced and made available for evaluation. He writes: "The critique which brings the undiscussed into discussion, the unformulated into formulation, has as the condition of its possibility objective crisis, which, in breaking the immediate fit between the subjective structures and the objective structures, destroys self-evidence practically" (1977, 169). Theorists have thus offered studies of how social

---

partial constituents of the psychological entity or mechanisms that constitute an individual's habitus. But since schemas for Haslanger can also be "embodied in individuals as a shared cluster of open-ended dispositions," such that what is in an individual's psychology has a "counterpart" in the wider culture (2012, 415), we might equally understand implicit biases to be a part of an individual or sociocognitive schema. Elsewhere, Haslanger suggests that Bourdieusian habitus is similar to her concept of "ideology," that is, the conscious and nonconscious shared discursive framework of representation that we use to understand the world (2012, 18). She does not state exactly how schemas and ideology are related, but I take it that ideologies are sets of interrelated schemas distinguished by the way they function to maintain the status quo (cf. Haslanger 2017). If this is correct, then ideologies are also intersubjectively virtual objects that differ from my understanding here of habitus as a psychological object.

structures can function to deposit in individuals a "habitus of resistance" (Clarke 2000, Medina 2013), or a "radical habitus," that is, an "amplified and politicized . . . habitual self-interrogation" (Crossley 2003, 55). Nick Crossley, arguing that radical habitus is formed in an ongoing way, even outside specific moments of crisis, writes:

> The individual acquires specific reflexive schemas for inspecting and defining their actions, perceptions, thoughts and feelings, and elects to work upon them to bring them into line with their new ideals. . . . They seek social change, in part, through self-change. Their activism entails an ongoing attempt to change their habitual ways of being-in-the-world: that is, "habit-busting habits." (56)

Insofar as the cultivation of radical habitus is crucial for social movements, then, at least *this* kind of modification to individual attitude is necessary for structural transformation (cf. Madva 2016). Indeed, self-change of this sort can *constitute* a form of structural transformation, where the type of structure being transformed is habitus.

I think responding to implicit bias is an important (though certainly not the only) way in which habitus gets radicalized. I thus defend a certain kind of standard story about implicit bias, adapted from the one Haslanger rejects:

> I am an employer who is considering three candidates for a job: Kwame, Kathy, and Eric. I am committed to treating all candidates equally but I have implicit gender and racial biases and although Kwame and Kathy have comparable strengths and weaknesses to Eric's, Eric appears to me to be the strongest candidate because he is a white male (though this latter fact is unbeknownst to me). Repeat this scenario—including cases of applications for educational opportunities, access to health and financial resources, etc.—and this provides part of an explanation of social inequality along lines of race/sex. (2015, 3)

This is a story of what Jennifer Saul (2013b) calls "bias-related doubt." The key difference between Haslanger's original story and this one is a shift from the third-person to first-person perspective. I think there is something very valuable about this story, mainly because it is not really a very standard story at all. As I have argued, it actually *is* a story about structures, albeit of habitus rather than traditional macro-level social structures. And it can trigger a moment of crisis: for *why* is it that I and so many others keep choosing the White male candidate over the female or non-White ones even when we are committed to equal treatment? What explains this lack of fit? Knowing my own sincere egalitarian commitments, and

those of many others, such a choice cannot be explained by my own moral agency. I am thus forced by such a story to bring to the surface precisely those "offstage" social structures that Haslanger (2015, 10) wants us to focus on—only I find them to be in my own head.

But it is precisely because it has the form of a standard story that the implicit bias story is able to generate this sense of normative urgency that other structural explanations lack. For I *qua* agent necessarily see myself as having freedom and agency whenever I take the first-personal practical perspective. From *this* perspective—not the third-personal explanatory perspective of the social theorist—my implicit biases appear to me as external forces, as structural constraints on my rational freedom, and hence as a crisis to be dealt with. While a story about large macro-level structures may not seem to implicate me or leave me any role to play, a story about structures in my own head cannot but involve me. This, I think, is the story that philosophers of implicit bias have been telling. It is a story that impels me to act, because action always begins from some particular individual's first-person perspective. As Saul (2013b, 243) writes in her account of bias-related doubt: "We feel perfectly fine about setting aside [traditional skeptical] doubts we have felt when we leave the philosophy seminar room. But with bias-related doubt, we don't feel fine about this at all. We feel a need to *do something* to improve our epistemic situation."

Haslanger's real worry about implicit bias seems to be that, in acting, I will try *merely* to correct my own biases. It is true that there is danger in putting too much store in individual "de-biasing" techniques and thinking that that is all justice requires. But no philosopher working on implicit biases defends such a claim. Again, the very nature of implicit biases—their invisibility, the fact that they are like water—generates the ever-present lurking possibility motivating the skepticism "that we *really should not* trust ourselves as inquirers" (Saul 2013b, 253). Such skepticism makes it clear that remedies for injustice cannot consist in individuals exercising their agency, but rather in the construction of other external structures that can undergird better internal structures.

Indeed, Haslanger herself acknowledges a number of practical advantages of discussing implicit bias, including as a starting point for discussion (because we all have biases), as a site for moral improvement, and as an explanation for the intractability of social injustice (2015, 12). But I hardly think that these are accidental features; rather, they derive from the structural nature of implicit bias. Understanding them as such thus creates more room for theorizing implicit bias than the "small space" that Haslanger allots (11). By directing our attention to specific junctures between structures where bias makes a difference (as in our everyday practices of grading, hiring, and peer-reviewing), efforts to address implicit bias represent attempts to intervene surgically on particularly vulnerable sites

within the complex and otherwise unwieldy edifice of social structure. Bourdieu writes of "identifying true sites of freedom, and thus of building small-scale, modest, practical morals in keeping with the scope of human freedom which, in my opinion, is not that large" (Bourdieu and Wacquant 1992, 199). More optimistically, I would say that by cultivating attentiveness to implicit bias, we begin to cultivate the kind of radical habitus that sustains and is sustained by social movements—in other words, *we begin to construct new social structures*, from the inside out. After all, most of the recommendations made by philosophers working on implicit bias have in fact been institutional and collective, for example, encouraging policies of anonymous review at philosophy journals, writing and disseminating good practice guides, and starting the Gendered Conference Campaign.[28] In trying to mitigate and block bias, we immediately see specific ways the social world must change for this to be successful, beginning with the small-scale structures over which we *do* have control.[29]

From the theoretical perspective of explaining social phenomena, then, Haslanger may be correct that we lack important kinds of autonomy and self-determination. But from the practical perspective of acting—of choosing between doing nothing, on the one hand, and seeking out others to work together, on the other—I *am* an individual who experiences the autonomy to choose actions (from the options set and constrained by my social milieu, to be sure) that constitute taking responsibility for structural injustice. Thus it is through storytelling and engaging one another as particular individuals that we mobilize and hold one another responsible for doing so. This, I conjecture, is why the first-person story of implicit bias and bias-related doubt has had such success in motivating many philosophers to be concerned about injustice within the discipline, and in spurring just the sort of collective action—the beginnings of new social structures—that we have seen in the wake of such a concern.

VI. Conclusion

I have argued that we need a theory of individual responsibility for structural injustice, understood as accountability rather than attributability, and I have argued

---

[28] The British Philosophical Association (BPA) and Society for Women in Philosophy (UK) Good Practices Scheme features "gender bias" as one of its five main categories. The Gendered Conference Campaign, run by the Feminist Philosophers blog, argues that all-male line-ups contribute to stereotyping and implicit bias against women in philosophy. Notably, one of the primary movers behind these efforts is Jennifer Saul, whose early work on implicit bias was largely responsible for bringing it to the attention of the discipline.

[29] I am grateful to an anonymous reviewer for this way of putting the point.

that individualistic practices of accountability are needed to address the problems of mobilization and solidarity inherent in all collective organizing. Additionally, I have proposed an alternative social-theoretic understanding of implicit biases as a type of Bourdieusian structure. In short, we can and should understand individuals as responsible (in the accountability sense) for structural injustice. And we can and should understand the effort to block and eliminate implicit biases as itself a kind of structural transformation.

References

Ahmad, Asam. 2015. "A Note on Call-Out Culture." *Briarpatch Magazine*, March 2. http://briarpatchmagazine.com/articles/view/a-note-on-call-out-culture.

Banks, Andy, and Jack Metzgar. 1989. "Participating in Management: Union Organizing on a New Terrain." *Labor Research Review* 8 (2): 1–58.

Berenstain, Nora. 2016. "Epistemic Exploitation." *Ergo, an Open Access Journal of Philosophy* 3 (22): 569–590.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

———. 1990. *The Logic of Practice*. Palo Alto, CA: Stanford University Press.

Bourdieu, Pierre, and Loïc Wacquant. 1992. *An Invitation to Reflexive Sociology*. Chicago: University of Chicago Press.

Brownstein, Michael. 2015. "Attributionism and Moral Responsibility for Implicit Bias." *Review of Philosophy and Psychology* 7 (4): 765–786.

Brownstein, Michael, and Jennifer Saul. 2016. *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*. Oxford: Oxford University Press.

Card, Claudia. 2010. *The Unnatural Lottery: Character and Moral Luck*. Philadelphia: Temple University Press.

Clarke, Melissa. 2000. "Rosa Parks' Performativity, Habitus, and Ability to Play the Game." *Philosophy Today* 44:160–168.

Clawson, Dan. 2003. *The Next Upsurge: Labor and the New Social Movements*. Ithaca, NY: ILR Press.

Coates, Justin, and Neal Tognazzini. 2013. *Blame: Its Nature and Norms*. New York: Oxford University Press.

Cobble, Dorothy S. 2007. *The Sex of Class: Women Transforming American Labor*. Ithaca, NY: Cornell University Press.

Collins, Patricia Hill. (1990) 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Cooper, Brittney. 2014. "It's Not About You, White Liberals: Why Attacks on Radical People of Color Are So Misguided." *Salon*, April 8, 2014.

http://www.salon.com/2014/04/08/its_not_about_you_white_liberals_why _attacks_on_radical_people_of_color_are_so_misguided.

Cross, Katherine. 2014. "Words, Words, Words: On Toxicity and Abuse in Online Activism." *Nuclear Unicorn* (blog), January 3. http://quinnae.com/2014/01/03/words-words-words-on-toxicity-and-abuse-in-online-activism/.

Crossley, Nick. 2003. "From Reproduction to Transformation." *Theory, Culture & Society* 20 (6): 43–68.

Dahrendorf, Ralf. 1968. *Essays in the Theory of Society*. London: Routledge.

Doan, Michael. 2016. "Responsibility for Collective Inaction and the Knowledge Condition." *Social Epistemology 30* (5–6): 532–554.

Fletcher, Bill, Jr., and Fernando Gapasin. 2008. *Solidarity Divided: The Crisis in Organized Labor and a New Path Toward Social Justice*. Berkeley: University of California Press.

Frankfurt, Harry. 1988. "Identification and Externality." In *The Importance of What We Care About: Philosophical Essays*, 58–68. Cambridge: Cambridge University Press.

Frye, Marilyn. 1983. *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, NY: Crossing Press.

Gendler, Tamar Szabó. 2011. "On the Epistemic Costs of Implicit Bias." *Philosophical Studies* 156 (1): 33–63.

Goodin, Robert E. 1987. "Apportioning Responsibilities." *Law and Philosophy* 6 (2): 167–185.

Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes." *Psychological Review* 102 (1): 4–27.

Haslanger, Sally. 2004. "Oppressions Racial and Other." In *Racism in Mind*, edited by Mark Levine and Tamas Pataki, 97–123. Ithaca, NY: Cornell University Press.

———. 2012. *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.

———. 2015. "Distinguished Lecture: Social Structure, Narrative and Explanation." *Canadian Journal of Philosophy* 45 (1): 1–15.

———. 2017. "Racism, Ideology, and Social Movements." *Res Philosophica* 94 (1): 1–22.

Held, Virginia. 1970. "Can a Random Collection of Individuals Be Morally Responsible?" *Journal of Philosophy* 67 (14): 471–481.

Holroyd, Jules. 2012. "Responsibility for Implicit Bias." *Journal of Social Philosophy* 43 (3): 274–306.

———. 2015. "Implicit Bias, Awareness and Imperfect Cognitions." *Consciousness and Cognition* 33: 511–523.

hooks, bell. 1984. *Feminist Theory from Margin to Center*. Boston: South End Press.

Isaacs, Tracy. 2011. *Moral Responsibility in Collective Contexts*. New York: Oxford University Press.

Jones, Robert P., Daniel Cox, and Juhem Navarro-Rivera. 2014. *A Shifting Landscape: A Decade of Change in American Attitudes about Same-sex Marriage and LGBT Issues*. Washington, DC: Public Religion Research Institute.

Kutz, Christopher. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Lavin, Chad. 2008. *The Politics of Responsibility*. Urbana: University of Illinois Press.

Lee, Frances. 2017. "Why I've Started to Fear My Fellow Social Justice Activists." *YES! Magazine*, October 13. http://www.yesmagazine.org/people-power/why-ive-started-to-fear-my-fellow-social-justice-activists-20171013.

Levy, Neil. 2014a. "Consciousness, Implicit Attitudes and Moral Responsibility." *Noûs* 48 (1): 21–40.

———. 2014b. "Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements." *Noûs* 49 (4): 800–823.

Lizardo, Omar. 2004. "The Cognitive Origins of Bourdieu's Habitus." *Journal for the Theory of Social Behaviour* 34 (4): 375–401.

Lorde, Audre. (1988) 2017. *A Burst of Light: And Other Essays*. Mineola, NY: Ixia Press.

Madva, Alex. 2016. "A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy." *Ergo, an Open Access Journal of Philosophy* 3 (27): 701–728.

Madva, Alex, and Michael Brownstein. 2016. "Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind." *Noûs*. doi:10.1111/nous.12182.

May, Larry. 1990. "Symposia Papers: Collective Inaction and Shared Responsibility." *Noûs* 24 (2): 269–277.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. New York: Oxford University Press.

Miller, David. 2001. "Distributing Responsibilities." *Journal of Political Philosophy* 9 (4): 453–471.

Parsons, Talcott. (1951) 1991. *The Social System*. London: Routledge.

Ransby, Barbara. 2003. *Ella Baker and the Black Freedom Movement: A Radical Democratic Vision*. Chapel Hill: University of North Carolina Press.

Saul, Jennifer. 2013a. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy: What Needs to Change?*, edited by Katrina Hutchison and Fiona Jenkins, 39–60. Oxford: Oxford University Press.

———. 2013b. "Scepticism and Implicit Bias." *Disputatio* 5 (37): 243–263.

Sewell, William H., Jr. 1992. "A Theory of Structure: Duality, Agency, and Transformation." *American Journal of Sociology* 98 (1): 1–29.

Springer, Elise. 2013. *Communicating Moral Concern: An Ethics of Critical Responsiveness*. Cambridge, MA: MIT Press.

Trần, Ngọc Loan. 2016. "Calling IN: A Less Disposable Way of Holding Each Other Accountable." In *The Solidarity Struggle: How People of Color Succeed and Fail at Showing Up for Each Other in the Fight for Freedom*, edited by Mia McKenzie, 59–63. Oakland, CA: BGD Press.

Wacquant, Loïc. 2005. "Habitus." In *International Encyclopedia of Economic Sociology*, edited by Jens Beckert and Milan Zafirovski, 315–319. London: Routledge.

Weingarten, Randi, Antonia Cortese, and Loretta Johnson. 2009. *Powering Up: The AFT Organizing and Union-Building Model.* Washington, DC: American Federation of Teachers, AFL-CIO.

Young, Cate. 2014. "A Refresher Course on Tone Policing." *BattyMamzelle*, April 11, 2014. http://www.cate-young.com/battymamzelle/2014/04/A-Refresher-Course-On-Tone-Policing.html.

Young, Iris Marion. 2011. *Responsibility for Justice*. New York: Oxford University Press.

Zheng, Robin. 2016. "Attributability, Accountability, and Implicit Bias." In *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, edited by Michael Brownstein and Jennifer Saul, 62–89. Oxford: Oxford University Press.

ROBIN ZHENG is an assistant professor of philosophy at Yale-NUS College. Her research focuses on issues of moral responsibility and structural injustice, along with other topics in ethics, moral psychology, feminist and social philosophy, and philosophy of race. Recent publications include "Why Yellow Fever Isn't Flattering: A Case Against Racial Fetish" in the *Journal of the American Philosophical Association*, "Precarity is a Feminist Issue: Gender and Contingent Labor in the Academy" in *Hypatia: A Journal of Feminist Philosophy*, and "Attributability, Accountability, and Implicit Bias" in *Implicit Bias and Philosophy: Volume 2* (edited by Michael Brownstein and Jennifer Saul). She has recently served on the APA Task Force on Diversity and Inclusion as well as the APA Task Force on a Good Practices Guide.