Valerie Soon

Penultimate draft – please do not cite. Final version available here:

https://link.springer.com/article/10.1007/s11098-019-01288-y

Implicit bias and social schema: a transactive memory approach

Abstract

To what extent should we focus on implicit bias in order to eradicate persistent social injustice? Structural prioritizers argue that we should focus less on individual minds than on unjust social structures, while equal prioritizers think that both are equally important. This article introduces the framework of transactive memory into the debate to defend the equal priority view. The transactive memory framework helps us see how structure can emerge from individual interactions as an irreducibly social product. If this is right, then debiasing interventions are structural interventions. One upshot is that the utility of the individual versus structural distinction is not apparent for the purposes of intervention.

1. Introduction

Formal equality is the law of the land in the United States, yet substantive equality in important respects remains elusive. Racial and gender inequality persist in many domains. Women and people of color tend to bump up against a glass ceiling at certain levels of management in the corporate world (Hyun 2009). There is now an extensive literature showing that medical professionals do not take the complaints of women and African Americans as seriously as those of white men, resulting in severe problems going undetected until it is too late (Hoffman & Tarzian 2001, Trawalter et al. 2012). And recent police shootings of African

Americans have raised the racial disparity in police brutality to the national consciousness: black Americans are 2.5 times as likely as whites to be killed by the police (Lowery 2016). A few decades ago, one could easily appeal to prejudice to explain these unjust disparities. But even with the recent resurgence of white nationalism and xenophobia in the United States and Europe, we are still hard-pressed to explain these disparities, since most people do not hold such extreme views. In fact, most people now claim that they not only do not hold prejudicial attitudes toward women and people of color, but find them abhorrent. Yet these same people might cross to the other side of the street every time they see a young black man walking their way, or they might dismiss a female colleague's suggestion as foolish while lauding a male colleague for making the same suggestion moments later. What explains these phenomena, given a significant reduction in legal discrimination and explicit prejudice?

Answers to this question fall into roughly two camps, individualism and structuralism. More will be said shortly about the distinctions within these two views. For now, the individualist stance can roughly be characterized as emphasizing the importance of individual minds. Structuralists, however, think that if structures are the root cause of injustice, biased minds are merely symptomatic of a deeper problem, so we should focus on structures rather than individuals. The focus of this paper is on a narrower disagreement within this debate: structural prioritism versus equal prioritism. Even if it is necessary to consider individual minds, we should still focus mainly on structures (Haslanger 2012, 2016). Call this position *structural prioritism*. Recent work in this literature, such as Machery et al. (2010), Madva (2016) and Davidson and Kelly (2018), has sought to undermine the utility of this distinction by highlighting the explanatory power of psychologistic explanations. Call this position *equal prioritism*. In this vein, this paper introduces the framework of transactive memory into the debate to defend a

focus on individual minds, but without relinquishing the importance of structures in shaping those minds. According to the transactive memory framework, a collective memory product can emerge from the interactions between individuals. This framework helps us see structure and individuals as two components of a dynamical system, guiding us to diagnose more effective sites of intervention. Individual minds are not merely shaped by structures, but actively shape structures. The implication is that implicit bias interventions are not merely necessary for social justice, but equally important as structural interventions.

It is important to note that this strategy departs from the state-based conception of the individualist-structuralist debate. The debate is currently founded on a conceptual distinction between minds and structures. On this conception, the question is: Are minds *or* structures the main ongoing cause of the inequality in question? As Davidson and Kelly (2018) have argued, this conceptual distinction is not a clean ontological one, because some informal "soft structures", such as norms, are internalized. My account is motivated by this ontological overlap to shift the debate toward thinking about the *processes* that create and sustain minds and structure. In other words, the focus is on the causal connection between the two rather than the causal weight of each element, considered in isolation, in producing injustice. By reframing the individualist-structuralist debate to emphasize the dynamic processes by which structures and biased minds mutually sustain themselves, we can see that our policy options do not take the form of either-structural-or-individualistic choices. Individualistic interventions can have structural effects, and vice versa.

I proceed as follows. In Section 2, I situate structural prioritism and equal prioritism within the conceptual landscape of the individualism-structuralism debate. I highlight the connection between explanation and intervention, as well as points of agreement between these

moderate versions. In Section 3, I leverage these points of agreement to arrive at the conclusion that structural interventions should not be prioritized over individualistic ones. I do so by drawing parallels between the processes of transactive memory and implicit bias: interactions between biased minds result in a memory schema, which just is one key component of a social schema, a quintessential element of structure. I conclude in Section 4 that we should therefore see debiasing as a structural intervention. This resolution undermines the utility of the distinction between structural and individualistic interventions on social injustice.

2. Varieties of individualistic and structural explanation

It is necessary to first clarify where structural prioritism and equal prioritism are situated in the contours of the broader individualism-structuralism debate. Within this narrower focus, two claims are of particular importance. First, both sides believe that the appropriate type of intervention on injustice follows from the correct type of explanation. Second, both sides think that ongoing, mutual feedback loops between individuals and structures are crucial for their respective arguments.

2.1. Individualism versus structuralism, broadly

One dominant explanation for persistent social injustice is psychological: it is caused by what's in individuals' hearts or heads (Garcia 1996, Blum 2002). Implicit bias clearly fits into this category. Implicit bias theories hold that "actors do not always have conscious, intentional control over the processes of social perception, impression formation, and judgment that motivate their actions" (Greenwald & Krieger 2006, 946). The Implicit Association Test (IAT), the most well-known measure of implicit bias, reveals associations between stigmatized groups of people and judgments. The more associations one has between stigmatized groups and

negative judgments, the higher one's implicit bias score. There is often dissociation between implicit bias and explicit evaluation on self-report measures: someone who has a low level of explicit bias might have a high level of implicit bias, either because they are unable or unwilling to report their bias (Nosek et al. 2002). It is important to note that the IAT does not pick up on distinctively unconscious states or processes, so there are multiple possible explanations for this dissociation (Hahn and Gawronski 2018). One explanation is that people have no introspective access to attitudes that cause implicit bias (Hofmann et al. 2009). But people could be unwilling to report these biases (Gawronski et al. 2007), or implicit bias might be experienced as spontaneous affective reactions that people reject (Fazio 2007, Gawronski & Bodenhausen 2006, 2011). Despite these distinctions, the general public's understanding of implicit bias is that it is unconscious, broadly speaking.

This position has received considerable uptake in the public sphere. According to the "public view", the mental states of individuals ultimately explain the persistence of injustice. It is hard to find a clear characterization of the public view, but the view seems to be that implicit bias is both necessary and sufficient for maintaining injustice, so we should focus our efforts on implicit bias interventions. Here are some examples of the public view in action: in a presidential debate during the 2016 election, Democratic Party candidate Hillary Clinton told Americans that "implicit bias is a problem for everyone, not just police." Implicit bias and other prejudice-reduction trainings (also known as "diversity trainings") are now de rigueur in workplaces ranging from police departments in Chapel Hill and Albuquerque to tech companies in Silicon Valley (Huet 2016). These trainings usually take the form of raising awareness of one's thought processes in order to avoid acting on stereotypical judgments (NC Racial Justice 2015). Debiasing procedures include counterstereotype training, in which subjects repeatedly affirm

non-stereotypical traits for certain groups. Kawakami et al. (2007), for instance, asked subjects to select in a given pair of traits the trait that was not culturally associated with the relevant group. For example, when subjects were presented with a photograph of a woman, the correct choice was "strong" as opposed to "sensitive" (Kawakami et al. 2007, 143). The idea is that retraining associations will result in a change in attitudes, beliefs, or behavior, and that this will gradually eliminate injustice. According to the public view, structures such as social norms, institutions and practices either reduce to mental states or are upheld primarily by mental states; injustice lies in the minds of individuals, so in order to eradicate injustice, we should change individual minds. This crude characterization of the public view finds precedence in a strong version of methodological individualism: "the doctrine that all social phenomena (their structure and their change) are in principle explicable only in terms of individuals – their properties, goals, and beliefs" (Elster 1982, 453; 2007)[1]. So for simplicity's sake, we can consider the public view and its strong methodological individualist foundations to constitute the *individualist view*.Though this is an important and influential view, it is not the view that I am most concerned with in this paper.

Some legal scholars and philosophers have criticized the individualist view as losing sight of the most important causes of inequality. According to this criticism, the implicit bias explanation is too focused on "biased minds" rather than social structural factors such as media representation, de facto segregation, unjust social relations (Haslanger 2012, 2015), and social norms and practices (Witt 2011). Call these critics, broadly, *structuralists*. Structuralists endorse

[1] Elster's methodological individualism hinges on the assumption that more fine-grained explanations are better explanations, so psychological explanations are the best social explanations. Many individualists do not accept this assumption (see List and Spiekermann 2013 for an overview of distinctions within individualism). See also Bratman 1993, 2014, Pettit 1993 for varieties of psychologistic individualism: the idea that social phenomena, such as group attitudes, are analyzable in terms of individual attitudes.

the idea that minds are shaped by the environment, so it would be inefficient and ineffective to debias individual minds without a corresponding change in the social environment. At worst, it would be actively harmful to do so: individualistic efforts would draw attention away from more lasting interventions, perhaps by luring us into political complacency, so a focus on individual psychology should be avoided. Banks and Ford (2009, 1054) exemplify this strong structuralist position:

> Despite its ostensible political benefits, the unconscious bias discourse is as likely to subvert as to further the cause of racial justice. Racial injustice inheres in the entrenched substantive racial inequalities that pervade our society. These disparities are not primarily a consequence of contemporary racial bias. Thus, the goal of racial justice efforts should be the alleviation of substantial racial inequalities, not the eradication of unconscious bias.

2.2. Structural priority versus equal priority

According to the general structuralist criticism of individualism, even if our biases did cause some injustice, they are only a proximate cause. Our associations between traits and groups are shaped by background structures, so structures are more explanatorily powerful. Strong structuralists take this as a reason not to focus on biased minds at all. *Structural prioritizers*, by contrast, think that though attention to individuals is necessary, structures should be our main point of focus due to their explanatory power (Haslanger 2016b).

Structural explanations explain "the behavior of a thing by explaining the behavior of something of which it is a part, if it is a part whose behavior is constrained by other parts of the whole" (Haslanger 2016a, 115). Individuals are part of social structures, so injustice is best

explained by reference to properties of social structures: a general category of social phenomena including "social institutions, social practices and conventions, social roles, social hierarchies, social locations or geographies, and the like" (Haslanger 2012, 413). Structures can also take the form of informal "networks of social relations" that are constituted through social practices (Haslanger 2016a, 125).

To illustrate the difference, consider the following explanations for the same social phenomenon (example borrowed and modified from Haslanger 2016a, 123-124, originally from Cudd 2006):

Individual: Lisa quit her job because she thought it would be best for her family.

Structural: Lisa quit her job because Larry makes more money, and Lisa is a woman who occupies the wife/mother node in a problematic structure of family/work relations.

What kind of additional information does a structural explanation provide? It gives us information about the background conditions that enable or constrain Lisa's choices. In other words, this is causal information.[2] Following Dretske's (1988) distinction between triggering and structuring causes, Haslanger argues that structural explanations give information about structuring causes, the causes responsible for a process being *this* particular process. By contrast, a triggering cause is responsible for the occurrence of a process now. Structuring causes are explanatorily prior to triggering causes. By focusing on structuring instead of triggering causes – background conditions instead of mental states – structural prioritizers hold that we are better able to explain stable patterns of social phenomena (Haslanger 2016a, 119), such as the general tendency of women rather than men to quit their jobs once they start families. If we want to

[2] As a reviewer pointed out, explanations need not be causal. However, it would be difficult to understand the tight connection between explanation and intervention in this debate without a causal notion of explanation (cf. Woodward 2003).

intervene on this pattern, we should target the structure of family/work relations rather than the preferences or other mental states of individual women.

Somewhere between individualism and structural priority lies the "anti-anti-individualist" position (Madva 2016). On this view, a focus on psychologistic interventions should be given *equal priority* as structural interventions. Equal prioritizers argue that individual change is necessary for any lasting structural change and should not be neglected (Davidson and Kelly 2018). It is implausible that debiasing would have no effect on inequality, and many social injustices, such as unequal death penalty sentences for whites and blacks, can be best explained by biased minds (Madva 2017, Cholbi and Madva 2018, Machery et al. 2010). According to implicit bias explanations, prejudiced individual attitudes and beliefs are a necessary part of an adequate explanation for injustice. This position implies that individual change is necessary, though not sufficient; we need to eradicate prejudiced psychological states in order to create a more just world, not just intervene on structures. Moreover, equal prioritizers also believe that individual change does not preclude a focus on structures. Machery et al, for instance, think that good psychological explanations will situate biased minds in the social environment and "show how racist thought and evaluation operate in context" (2010, 243).

Madva thus sums up equal prioritism:

> Ultimately, the attempt to force a choice between individual and structural change is confused. My view is not that we should instead prioritize individual change, but that individual changes will be integral to the success of structural changes. (Madva 2016, 702)

2.3. From explanation to intervention

Structural prioritizers and equal prioritizers both agree that structures and individuals are both necessary, and neither alone is sufficient, for social change to occur. This agreement makes them more nuanced and moderate views in the conceptual landscape. However, the core disagreements that motivate the individualism-structuralism debate still remain. The structural prioritizer still clearly throws her hat in with hardcore structuralists such as Banks and Ford (2009). Given the nature of the dialectic, the equal prioritizer's insistence on the importance of individual minds is more closely aligned with the individualist position, despite occupying the middle ground as a purely substantive matter.

Most obviously, the first point of disagreement between individual prioritizers and structural prioritizers centers on *explanatory priority*: are mental states or structures the best, i.e. more deeply entrenched or robust, explanation of persistent social injustice? Note that neither side denies the necessity of the other type of explanation, merely its priority.

A second related point of disagreement has to do with *priority of interventions*. Though the type of explanation does not entail the type of intervention, the connection between the two is very close. Prioritizing either structures or individuals in an explanation is supposed to tell us where to target our limited resources for intervention. Thus, the type of explanation proffered has implications for whether an intervention will be effective in ameliorating social injustice, and whether we are focusing our resources on the right target. Due to the explanatory priority of structures, structural prioritizers do not think we should prioritize individualistic interventions such as debiasing. Haslanger writes, "If the best explanation of social stratification is structural, then implicit bias seems at best tangential to what is needed to achieve justice" (2015, 2). The worry for the structural prioritizer is that attempts to rehabilitate the mental states of particular individuals, e.g. through bias reduction training, are merely palliative if undertaken without

attention to how these mental states are shaped and consistently affirmed by social structures. As Haslanger puts it, "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" (2015, 10). Thus, the public focus on individual-level reforms such as implicit bias reduction is misguided. Psychological individualistic explanations are ultimately ineffective and insufficient at best because they don't target the background causes of injustice, and consequently prescribe the wrong type of intervention. If we are to focus on implicit bias,

> an adequate account of how implicit bias functions must situate it within a broader theory of social structures and structural injustice; changing structures is often a precondition for changing patterns of thought and action and is certainly required for durable change. (Haslanger 2015, 1)

One might object that these are really two distinct discussions that are too often conflated: we can talk separately about the best explanation for injustice and the effectiveness of intervention. For example, discrimination that is straightforwardly caused by individuals might best be targeted by a structural intervention, such as antidiscrimination law. In comments on Madva (2016), Ayala has argued that Madva's defense of debiasing interventions rests on such a conflation. Nevertheless, as some of the statements above show, the concepts of explanation and intervention are tightly bound together, rather than conflated, in this debate. Both sides take the predicted effectiveness of intervention to be an upshot of the correct type of explanation for the phenomena.

Structural prioritizers also marshal the resources of cognitive science to support the connection between explanatory priority and priority of intervention. Huebner (2016a), for example, argues that implicit bias is the result of scaffolded moral cognition. According to

Sterelny's (2010) scaffolded cognition hypothesis, some of our cognitive processes could not be carried out without environmental resources. We then shape the environment to facilitate cognition. Huebner argues that our low-level (i.e. reflexive) cognitive systems are not just attuned to the social environment; they also rely on the social environment to generate optimal behavioral policies for the agent. For example, we take the emotional reactions of our social group, as well as representations of racialized out-group members, to signal something about environmental threats or rewards. We fall back on these lower-level systems to guide behavior when under cognitive load. Implicit biases, then, are just "situational adaptations that are attuned to features of the racist, sexist, and heteronormative communities in which we are immersed" (Huebner 2016a, 8, from Dasgupta 2013, 240). Huebner's scaffolded moral cognition hypothesis predicts that in the absence of an unjustly structured social environment, the products of our social cognition, such as associations between groups and concepts, would track different environmental regularities. Accordingly, Huebner thinks that the best way to intervene on implicit bias is to create a more egalitarian world instead of focusing on the biases themselves:

> So as we watch or read the news, watch films, rely on tacit assumptions about what is likely to happen in particular neighborhoods, or draw illicit inferences on the basis of the way in which a person is dresses, we cause ourselves to backslide into our implicit biases. No matter how calm, vigilant, and attentive to our biases we try to be, I maintain that we will be unable to moderate or suppress all our problematic implicit biases until we eliminate the conditions under which they arise. (Huebner 2016a, 71)

Because individual attitudes and actions are shaped in this way by structure, structural prioritizers contend that it is a mistake to focus on individual psychology if it is the structural background that guides individual behavior. For instance, Witt argues that we should shift "away

from a primary focus on individual psychologies, their gender schemas, deformed preferences, and unconscious biases, and toward the social world and its normative structure" (2011, 129). Focusing on individual psychology, by contrast, leaves the normative structure of the social world intact.

In short, to say that structural explanations are best is to say that they are more complete, and consequently more effectively guide interventions. Individualistic explanations are not wrong or false per se, but they can mislead us to focus on a site of intervention that is secondary in causing the injustice. By contrast, equal prioritizers think the priority of structures is not borne out by the evidence, and that individual-level factors interact with structural ones to perpetuate injustice. Thus, we should give as much priority to psychological as to structural interventions (Madva 2016, 712-13; Machery et al. 2010, 242-43; Davidson and Kelly 2018, 14-15).

1.3. Social schemas, feedback loops, and MIRROR

Let us now turn to some points of agreement between both sides. Both parties agree that the relationship between individuals and structures is characterized by a mutual, ongoing feedback loop. Structures constrain and enable our thoughts and behavior by making certain features of the world more salient than others, thereby shaping our attitudes and preferences (Haslanger 2016a, 128). We internalize structures in the head and subsequently act in accordance with them, perpetuating structures as a result. Consequently, structures are both in the head and out in the world as *social schemas*:

> clusters of culturally shared concepts, beliefs, and other attitudes that enable us to interpret and organize information and coordinate action, thought, and affect. Schemas are public—think of them as social meanings conventionally associated with things in our

social world, including language—but are also internalized and guide behavior. (Haslanger 2016a, 126)

At this point, one might raise the question: if there is a feedback loop between structures and individuals, why are structures explanatorily prior? One reason is that structures cannot be explained by mental states because they can persist even in the absence of mental states that align with structures. As Haslanger states, "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" (2015, 4). The simple tendency to follow norms and conventions can perpetuate unjust structures. For example, Ayala and Vasileyva argue that testimonial injustice can occur in the absence of prejudiced attitudes towards speakers from marginalized communities. Testimonial injustice is "rather the unfortunate result of perfectly skilled listeners who are appropriately applying the conventions operative in their communities" (2015, 3; also see Ayala 2018).

The structural prioritizer's conception of the feedback loop between individual and structure suggests the following view: our minds internalize social schemas and reflect them back out into reality. Madva calls this view MIRROR:

> ...to say that our biases are "mirror-like reflections" of the social world is to say that they are acquired and sustained simply by virtue of the fact that we grow up and remain immersed in a society structured by visible disparities between social groups. (Madva 2016, 717)

Reviewing a wide variety of psychological evidence, Madva rejects MIRROR as "radically oversimplified and misleading" (2016, 717-720). This is the key to his defense of equal prioritism: if MIRROR is false, then we should not prioritize structural interventions. Madva

points out that structural prioritizers seem to rely on MIRROR or it would be difficult to make sense of their arguments (2016b, 8-9), though elsewhere they reject MIRROR as too passive a view of the mind (Ayala 2016, Haslanger 2016b). In comments on Madva, Haslanger states that "the most promising model is one that puts *collaborative agency* [italics mine] at the center" (2016b, 3):

> On this approach, the mind does not passively "mirror" the world. Rather, we are engaged in complex cooperative and signaling practices that require cognitive coordination, and we teach each other what matters in order to participate with others. (Haslanger 2016b, 3)

This rejection of MIRROR, combined with the commitment to feedback loops, raises an internal tension for structural prioritizers. Can structural prioritizers explain how individuals can actively construct and transform structures while preserving structural prioritism? On a collaborative, interactive conception of the individual-structure relationship, individuals do not simply reflect internalized schemas back into the world. Individuals also construct and transform schemas together, such that schemas would be different if not for the minds and actions of individuals. As Huebner writes, we "simultaneously create and inhabit" (Huebner 2016a, 6) the environment, molding it with our patterns of behavior and expectations. Expectations shape our cognitive processes, which in turn lead us to respond in certain ways to the environment. Our responses subsequently shape the environment itself, influencing our own expectations as well as those of others. But if this is the case, why prioritize structural over debiasing interventions?

My aim in the next part of this essay is to answer this question by taking up Haslanger's suggestion of a model of collaborative agency. This suggestion seems promising, for as Doris (2015) argues, human agency is necessarily socially embedded. I leverage agreement about the

feedback loop between individuals and structures and the rejection of MIRROR to resolve the disagreement about whether individualistic or structural interventions should be prioritized. Using the framework of transactive memory, I sketch a model of collaborative agency that highlights the connection between biased minds and social schema. On this model, implicit bias both causes and partly constitutes a form of social schema. If this is right, there is no reason to accord priority to structural explanations and interventions. We should prioritize debiasing interventions as much as structural interventions.

3. A transactive account of implicit bias

What model of collaborative agency could deliver an account of the relationship between individuals and structures? Here, I draw an analogy between the process of implicit bias formation and transactive memory. I am not claiming that implicit bias is an instance of transactive memory. I simply aim to show that the collaborative aspect of transactive memory is instructive for thinking about how biased minds create structure. I first describe Wegner's (1987, 1991) theory of transactive memory, and argue that the process of transactive memory parallels the process of implicit bias formation in important ways. Interactions between individuals create and transform a memory schema, which is another way of describing one component of structure, a social schema. If this is right, structural prioritism is false.

3.1. Transactive memory: an overview

Transactive memory is the theory that a "set of individual memory systems in combination with the communication that takes place among individuals" creates a group product (Wegner 1987, 186). We scaffold cognition on other agents; by distributing information storage and cueing memories, they perform an active role in helping us remember. For example,

couples, close friends, and team members who spend a lot of time together often become adept at navigating complex problems, reconstructing experiences, and retrieving information together in a way that would not be possible with unfamiliar others. The model of transactive memory also scales up to contexts beyond those of close friendship and partnerships. Recently, it has been employed in organizational psychology to describe how firms encode, store, and retrieve information. If teams within organizations operate as transactive memory systems (TMS), then learning about how TMS works can help organizations function more efficiently (Ren & Argote 2011). These insights can also apply to social structure.

In a transactive memory system, individual memories are linked to a collective knowledge network through a division of labor (Ren & Argote 2011, 192-193). This is its structural component. A TMS also has a process component consisting of knowledge-relevant transactive processes such as encoding, storage, and retrieval. Successful encoding occurs when a commonly accessible label of some kind is attached to the memory content, and both the label and the content are internally stored within the mind of an individual. Then, transactive retrieval begins when one person asks another for a piece of information that she does not hold internally. Through communication, individuals involved in the retrieval process determine the location of the memory, which is internally stored under a label. Crucially, the memory label is not just a tag, but suggests a "common experience and provides a common foundation for explanation and elaboration". A label thus provides an "interpretive scheme" to organize different types of information under. It ties together disparate pieces of information internally stored by individuals and integrates them, creating a new piece of information upon retrieval. The collective memory consequently emerges as a result of the collaboration among individuals. A flaw in an individual link in the network can result in a collective memory that is inaccurate or distorted. Crucially, the

collaborative aspect of memory makes it a group product. It is not reducible to the mental operations or content of any individual within the group. This collective memory subsequently guides the beliefs and actions of individuals in the group (Wegner 1987, 187-197).

> To illustrate, consider the following organic case of transactive memory:
>
> Imagine, for example, that a couple is leaving a party. At different times, they each talked to Tex. The male notes that Tex was depressed this evening; he stared at the floor and barely talked. The female says that Tex was not at all depressed; in fact, she saw him for quite a while early in the party and he seemed unusually frisky and friendly. The male recalls that Tex said he was thinking about separating from his wife. And in short order, the couple reaches a conclusion: Tex was flirting with the female and embarrassed about it in the presence of the male. (Wegner et al. 1985, 267)

The theory has also been tested in the lab. Wegner et al. (1991) conducted a study on the memory performance of individuals who had been dating for at least three months. Natural couples performed better on recall tasks than assigned couples when they were not given an organizational scheme for performing the memory task. However, natural couples performed worse than assigned couples when an organizational scheme was imposed on them. Wegner et al. hypothesized that this discrepancy may occur because the process of assigning responsibility for knowledge in transactive memory systems works best when responsibility is implicitly assigned. An unnatural organizational scheme impairs memory by introducing uncertainty about the task or otherwise cognitively disrupting the flow of the memory performance (Wegner et al. 1991, 928). This conclusion also holds in teams and organizations, where team members who trained with a different group than the one they performed with did not develop TMS as strong as those who trained and performed in the same group (Moreland, Argote, and Krishnan 1996).

It is important to note that a *bidirectional flow of information* makes a memory system qualify as transactive. Individuals should be seen as a transactive memory system when they "cue, re-cue, and acknowledge one another's claims as they attempt to construct a plausible representation of what happened to them" (Huebner 2016b, 61). A unidirectional flow of information, by contrast, is an exploitative rather than a transactive relationship in which an individual uses some environmental resource as a tool to aid her cognitive processes, but does not shape that resource.

To sum up this brief overview, the dynamic interplay between structural and process components integrates the knowledge that each individual initially possesses, thereby generating new collective knowledge (Lewis & Herndon 2011, 1256). Importantly, the communicative process allows these individuals to generate the collective product; the product does not reduce to the aggregation of what's in individuals' heads.

3.2. Implicit bias: semantic associative memory and schemas

I now argue that the process component of transactive memory parallels the way that we encode, store, and retrieve the semantic associations involved in implicit bias. Though the parallels are not perfect, they illustrate how individual interactions generate a collective product. This is enough to get us the conclusion that implicit bias interventions are structural interventions. From this, we can infer that structural prioritism is false. Minds are not merely shaped by structure, but actively construct structure.

The prevailing view among philosophers and psychologists is that associative memory is a key process driving implicit bias (Huebner 2016b).[3] According to this view, implicit bias is the

[3] See Mandelbaum (2016) for an argument that implicit bias is not caused or constituted by semantic associations, but by unconscious propositionally structured beliefs. It is not within the scope of this paper to defend the associative theory of implicit bias, so I will simply say for now that the truth of my argument is conditional on the truth of the associative theory. Also see Del Pinal and Spaulding (2018)

result of a pattern of learning and unlearning where "semantic associations are formed across repeated stimulus pairings in a probabilistic fashion" (Amodio 2009, 8). When a particular concept is activated by an environmental stimulus, other semantically related concepts are also activated. In a series of implicit association tests conducted by Amodio and Devine, participants more frequently associated the attributes of athleticism, rhythmicity and unintelligence with African Americans (2006, 655). Amodio and Devine hypothesized that a semantic associative memory system is responsible for driving these results. One possible explanation for these results is that these attributes describe common and readily available depictions of African Americans in the cultural environment, leading to learned associations of African Americans with these concepts. When the concept of "African American" is activated, so are the associated attributes. These semantic associations constitute memory schemas, "existing knowledge structures which guide and facilitate the processing of social information" (Augoustinos & Innes 1990, 241). Similarities between semantic associations and memory schemas provide evidence of this. Memory schemas have the following features (Ghosh & Gilboa 2014). There is an *associative network structure* based on units and their relationships. Schemas are based on *multiple episodes*, so they are "higher-level constructs that encompass representations of the similarities or commonalities across multiple events" (Ghosh & Gilboa 2014, 106). Because no two episodes are identical, there is *lack of unit detail*. Finally, schemas are adaptable, meaning they are sensitive to assimilation and accommodation. Assimilation consists of "incorporating environmental elements into a schema without challenging the existing relationships in a

for the view that a conceptual dependency network model is better than an associative model for explaining how implicit biases are represented.

schema," while accommodation refers to "modifying a schema under pressures from new environmental elements" (Ghosh & Gilboa 2014, 108).

Finally, note that the associations between concepts and stimuli exhibit patterns of stability across individuals in a given environment. Stability means that certain concepts tend to reliably activate together among different individuals, as in the Amodio & Devine (2006) study. This holds regardless of individuals' other mental states or attributes. Further, associations are very difficult to unlearn (Amodio & Ratner 2009, 145). Such associations exhibit stability despite the fact that most individuals do not interact directly with each other. I return to this point in section 3.3.

The following example illustrates how semantic associative processes, combined with interaction among individuals, creates an irreducible group product. When an experience or knowledge is encoded, a label attaches to that content. For example, when one repeatedly hears political commentators or acquaintances make disparaging comments about professional football players who are protesting police brutality, one attaches the label "black football players" to the content "athletic individuals who aren't smart enough to make informed judgments about politics."[4] The association between stimulus (the label "black football players") and meaning (the content) is then stored in the individual's head. Imagine that this encoding and storage process occurs enough times among enough individuals. As a result, the label comes to imply "a common experience and provide a common foundation for explanation and elaboration" among agents (Wegner 1987, 195). It is an understanding that facilitates communication with others. The interactions between individuals, or between individuals and the environment, then stimulate

[4] The example here refers to the NFL protests, led by Colin Kaepernick, over police brutality. See Wilkinson (2018) for an overview of what responses to the protests reveal about political polarization.

retrieval of learned associations stored under the label. When another agent or the environment activates a label or one of the concepts falling under it, the first agent reflexively retrieves other concepts related to the first. Further communication with other individuals then constructs a shared memory product. For instance, one person might start out with an association between black football players and lack of patriotism, but no other associations. Another person might start out with an association between black football players and unintelligence, but not think that these football players were also unpatriotic. In communication with each other, these two people arrive at the group product: "Black football players are unpatriotic *because* they are unintelligent." The connection is not aggregative and does not reduce to what's in individuals' heads. Thus, it is an irreducibly social rather than individual product.

3.3. Memory schemas as social schemas

The final piece of my argument connects memory schemas and social schemas. Memory schemas are knowledge structures that facilitate the processing of information and subsequent action. I have argued that the semantic associations that constitute implicit bias are memory schemas, for they carry out precisely these two functions. Further, it is plausible that they are also social schemas. Recall what Haslanger says about social schemas and their relationship to structure:

> Social schemas consist in clusters of culturally shared concepts, beliefs, and other attitudes that enable us to interpret and organize information and coordinate action, thought, and affect. Schemas are public – think of them as social meanings conventionally associated with things in our social world, including language – but are also internalized and guide behavior. Both concepts and beliefs, in the sense intended,

> store information and are the basis for various behavioral and emotional dispositions (Haslanger 2016, 126).

A social schema, in other words, is a component of structure that does not reduce to individual mental states. Crucially, the memory schemas involved in implicit bias are "culturally shared" between individuals—they are stable across contexts and across diverse types of individuals. Stability means that certain concepts tend to reliably activate together among different individuals, as in the Amodio & Devine (2006) study. This holds regardless of individuals' other mental states or attributes. Further, associations are very difficult to unlearn (Amodio & Ratner 2009, 145). Such associations exhibit stability despite the fact that most individuals do not interact directly with each other. They are also internalized and guide behavior. Given these relevant similarities, why not think of memory schemas involved in implicit bias simply as social schemas? There are, of course, other social schemas that may not involve implicit bias, such as social norms (Ayala and Vasileyva 2015, Ayala 2018).

My claim is not that implicit bias schemas exhaust social schemas, simply that they are one important kind of schema. If memory schemas generated by the interactions among biased individuals are social schemas, the upshot is that debiasing interventions are not individualistic or a waste of resources. Because we tend to associate with those who hold similar ideological views to us (Bishop 2012), it is extremely likely that something like the above case happens every day. Transactive construction of semantic associations happens not only between close individuals in face-to-face interactions, but also online. Witness how easily "fake news" and harmful stereotypes propagate via social media disinformation campaigns (Bradshaw & Howard 2018). People not only spread false and harmful information, but also trust this information when it comes from select sources, and transform this information by affirming and adding to each

other's views, as if in a perverse game of telephone. Now that our media sources are significantly constituted by social networks rather than mass media, it is highly plausible that something like transactive memory is becoming more and more responsible for the production and propagation of culturally shared information such as stereotypes. Indeed, the transactive memory framework predicts something like the wildfire-like spread of misinformation. Information flows more easily, and collective memory products develop more easily, among closely knit groups than unfamiliar ones. For instance, transactive memory research on teams finds that familiarity between group members affects the development of TMS (Lewis 2004, He et al. 2007). Ren & Argote (2011) hypothesize that groups in which members are more familiar with each other will develop a TMS more easily than those in which members are less familiar with each other. That is, "members are more likely to divide areas of expertise, rely on each other, and share knowledge when they identify with the group than when they do not" (Ren & Argote 2011, 204). In order to nip this information-sharing in the bud, we must intervene on biased minds no less than on other components of the social environment.

3.4. Individual interventions and structural dynamics

The example I outlined above illustrates that debiasing can happen organically in everyday interactions. Individuals can influence their social milieu by retrieving positive associations and communicating those to their social group. But in order for debiasing to do the work I'm asking of it, it must have some effect on the social environment writ large. The relevant type of intervention must make individual change radiate out from the individual. Thus, for debiasing to come about in everyday interaction, there must be uptake by the other individuals in a group: biased individuals have to be receptive to new, positive associations. But

biased individuals are likely to only be receptive to certain types of persons, such as those in power or those who are close to them. The worry here is that debiasing interventions will be nullified by such communicative frictions, rendering them ineffective beyond the particular individual targeted.

One upshot of the transactive memory framework is that more attention must be paid to the structure of social networks, especially to the social position of individuals within these networks. How well-positioned are individuals to transmit and retrieve information within their groups? How likely is it that their words and actions will get taken up sincerely by others in their group, so that their input has influence on the final social product? Here, we can look to the literature on social norm change to see where we should focus on debiasing interventions. Bicchieri and Funcke (2018) have argued that social norm change requires "trendsetters" or "first movers" to lower the costs of deviating for everyone else. To be maximally efficient, debiasing interventions should focus on *certain types* of individuals: powerful insiders with social credibility[5], and autonomous individuals who are relatively insensitive to the pressures of social conformity.

Two examples illustrate this point. The recent #MeToo movement is a case of social norm change that was catalyzed by powerful insiders. Actress Alyssa Milano began encouraging use of the hashtag in response to allegations of sexual misconduct by Hollywood producer Harvey Weinstein. According to Twitter, the hashtag was tweeted over a million times in 48 hours; on Facebook, 45 percent of users had friends who posted "me too" (CBS News). The movement quickly sparked discussion about sexual harassment and reform in institutions such as

[5] There may also be a normative reason to focus debiasing efforts on powerful individuals – their social power may give them a duty to undermine unjust structures.

governments, finance, churches, and education. But the movement had actually begun over a decade earlier – community organizer Tarana Burke began using the phrase to bring attention to sexual abuse and assault, but did not receive the same uptake as Milano (Ohlheiser 2017). Why didn't Burke receive any of the attention that Milano did? A highly plausible explanation for this discrepancy is that Milano is a celebrity, as were many of the women who first began spreading the hashtag.

For an instance of norm change catalyzed by less powerful but autonomous individuals, we can look to the Civil Rights movement. Andrews and Biggs (2006) found that during the early Civil Rights Movement, there was a positive relationship between the number of autonomous[6] and affluent black adults and college students in a city, and the number of sit-ins in that city. Bicchieri hypothesizes that "Autonomous black adults had a greater capacity to dissent, and black college students likely perceived dissent to be less risky" because they had no jobs to lose (2017, 164). A focus on these two groups – powerful insiders and autonomous individuals – thus promises to help individual change radiate through social networks.

3.5. The individualist-structuralist distinction

What is the point of the explanatory distinction between individuals and structures? One purpose is to diagnose the site of intervention most likely to deliver durable change. On the transactive model I have sketched, the distinction between individuals and structures is not useful for this purpose, because individuals and structures independently influence each other. Individual interactions create a social schema, which then is internalized into individual heads, and so on in a feedback loop. According to dynamical systems theory, the presence of ongoing

[6] As inversely measured by "the percentage of the male labor force relegated to unskilled occupations – servants and laborers" (Andrews and Biggs 2006, 760).

mutual feedback loops signals that two seemingly distinct things really constitute one system (Palermos 2014). Neither variable can be conceived of simply as an input or output into the other; both variables are endogenously determined, and each variable performs an active, independent role in maintaining the feedback loop. Where these feedback loops are present, the division between two apparently distinct entities dissolves; intervening on one variable changes the dynamics of the whole system.

Though structural prioritizers acknowledge the mutual, ongoing feedback loop between individuals and structures, they have not done justice to this insight. Their emphasis has been on the influence of structures on individuals, portraying individuals as passive receptors of structural influence. The transactive memory framework, however, shifts the focus to the *independent* influence of individuals on structures. The arrangement of and interactions between individuals, in conjunction with the memory content in each of their heads, have the potential to either amplify or diminish the signal that structure sends out, thereby creating new structural elements, solidifying existing ones, or gradually eliminating them. This allows us to see the relationship between structure and individuals as a truly dynamic feedback loop in which both variables perform an active role in maintaining the system.[7] In other words, individuals are not just static variables that receive some causal impact from social structure, then feed this impact back in a loop. Individuals are agents whose actions have some causal effect on structure.

To sum up, the explanatory distinction between structures and individuals is supposed to tell us which site of intervention will be most fruitful. On the dynamic, interactionist picture offered by the transactive memory framework, there are reverberating influences between

[7] This approach has been employed by Skorburg (2017) in other debates. Skorburg argues that dynamical systems theory gives us an interactionist perspective that helps us dissolve dichotomies in the person-situation debate in social psychology and ethics (i.e. whether the person or their environment is more responsible for behavior), and the extended cognition literature.

structure and individuals. Structural prioritizers have focused on the effects of structures on individual minds, but I hope to have shown that the influence goes the other way as well: interventions on individual minds can have structural effects. The dichotomy between structures and individuals ignores this interplay, but in doing so, diminishes the importance of a key site of intervention: individuals, their organization, and interaction. With Machery et al. (2010), Madva (2016) and Davidson and Kelly (2018), I propose that we move away from this dichotomy. It distracts us from the goal of finding likely effective interventions.

4. Conclusion

I began by outlining the debate between equal prioritizers and structural prioritizers, culminating in an internal tension for the structural prioritizer: if she rejects MIRROR, the view that the mind is a passive reflection of reality, then what is the basis for prioritizing structural interventions over apparently individualistic debiasing interventions? And if she accepts the feedback loop picture of the relationship between individuals and structures, this seems to further undermine structural prioritism.

Using the theory of transactive memory, I have attempted to offer a model of collaborative agency that connects individuals to the environment, without loss of individual agency. Beginning with two premises friendly to both sides – rejection of MIRROR and commitment to feedback loops – I have argued that the individuals actively construct and maintain social schemas together, while also being guided and shaped by social schemas. Communication among individuals creates an irreducible group product, a memory schema. This is simply one component or one type of social schema. The upshot of this connection between bias and social schema, however, is that we should give up structural prioritism. As Madva has

argued (2016, 2017, 2018 with Cholbi), we should not deprioritize debiasing interventions. Debiasing interventions have the potential to deconstruct problematic social schemas. By eradicating bias within individual minds, we target the building blocks of negative semantic associations, and consequently of unjust social schemas.

What would such interventions look like? The equal prioritizer and structural prioritizer do not seem to disagree on this practical question, giving support to the idea that the importance of the individual-structural distinction is oversold. I have argued that debiasing interventions should focus on certain types of individuals in order to ensure that individual changes radiate throughout social networks. Debiasing interventions also need to be undertaken in conjunction with structural interventions in order to have maximal impact. If we want to deconstruct unjust social schemas, perhaps we should structure the social environment to promote interactions between diverse individuals so that different associations will be created (Anderson 2010). Huebner suggests that "we must attempt to reject dominant social norms, challenge existing social institutions, and develop practices that are better than those we have come to expect" (Huebner 2016b, 21). The hope is that these "collective prefigurative practices" will help bring our reflexive attitudes in line with our reflective ones (Huebner 2016b, 21). Holroyd and Kelly (2016) have also argued that one way of ameliorating implicit bias is to exercise "ecological control" over the environment, such that we enforce positive semantic associations and avoid triggering negative ones. In light of these suggestions, we can see that debiasing need not just take the form of implicit bias trainings in workplaces and schools, though these are crucial too. Institutional and material change are also debiasing methods. Structural prioritizers are no doubt correct that their methods of intervention on the material environment are needed to change biased minds. But equal prioritizers are also right about the importance of psychologistic

intervention for changing institutional and material structures. Not only are structural and individualistic interventions both essential; they are also not obviously distinct. For the sake of social justice, we should begin to move away from this binary.

References

Andrews, K. & Biggs, M. (2006). The dynamics of protest diffusion: movement organizations, social networks, and news media in the 1960 it-Iins. *American Sociological Review, 71*(5), 752-777.

Ayala, S. and Vasileyva, N. (2015). Explaining injustice in speech: individualistic vs. structural explanation. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, and P.P. Maglio (eds.) *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Ayala, S. (2016). Comments on Madva (2016). *Ergo* symposium.

Ayala, S. (2018). A structural explanation of injustice in conversations: it's about norms. *Pacific Philosophical Quarterly*, *99*(4), 726-748.

Amodio, D. & Devine, P. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*(4), 652.

Amodio, D. (2009). The social neuroscience of intergroup relations. *European Review of Social Psychology 19*(1), 1-54.

Amodio, D. & Ratner, K. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*(3), 143-148.

Augoustinos, M. and Innes, J. (1990). Towards an integration of social representations and social schema theory. *British Journal of Social Psychology*, *29*(3), 213-231.

Banks, R. & Ford, R. (2009). (How) does unconscious bias matter: law, politics, and racial inequality. *Emory Law Journal, 58*(5), 1053-1122.

Bicchieri, C. & Funcke, A. (2018). Norm change: trendsetters and social structure. *Social Research 85*(1), 1-21.

Blum, L. (2002). Racism: what it is and what it isn't. *Studies in Philosophy and Education*, *21*(3), 203-218.

Bradshaw, S. & Howard, P. (2018). Challenging truth and trust: a global inventory of organized social media manipulation. Working Paper Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk.

Bratman, M. E. (1993). Shared intention. *Ethics*, *104*(1), 97-113.

CBS News (October 17, 2017). More than 12M "Me Too" Facebook comments, posts, reactions in 24 hours. URL = <https://www.cbsnews.com/news/metoo-more-than-12-million-facebook-posts-comments-reactions-24-hours/>

Cholbi, M. & Madva, A. (2018). Black Lives Matter and the call for death penalty abolition. *Ethics* 128, 517-544.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.

Davidson, L. & Kelly, D. (2018). Minding the gap: bias, soft structures, and the double life of social norms. *Journal of Applied Philosophy*.

Del Pinal, G. & Spaulding, S. (2018). Conceptual centrality and implicit bias. *Mind & Language* 33(1), 95-111.

Doris, J. (2015). *Talking to ourselves: reflection, ignorance and agency*. Oxford University Press.

Elster, J. (1982). The case for methodological individualism. *Theory and Society*, 11(4), 453-482.

Elster, J. (2007). *Explaining social behavior: more nuts and bolts for the social sciences.* Cambridge University Press.

Fazio, R. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603-637.

Gabrieli, J. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*(1), 87-115.

Garcia, J. L. (1996). The heart of racism. *Journal of Social Philosophy*, *27*(1), 5-46.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us?: Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*(2), 181-193.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In *Advances in experimental social psychology* (Vol. 44, pp. 59-127). Academic Press.

Ghosh, V. and Gilboa, A. (2014). What is a memory schema? A historical perspective

on current neuroscience literature. *Neuropsychologia*, *53*, 104-114.

Greenwald, A.G. and Hamilton Krieger, L. (2006). Implicit bias: scientific foundations. *California Law Review 94*(4), 945-967.

Hahn, A. & Gawronski, B. (2018). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*.

Haslanger, Sally (2012). *Resisting Reality*. Oxford University Press.

Haslanger, Sally (2015). Social structure, narrative, and explanation. *Canadian Journal of Philosophy 45*(1), 1-15.

Haslanger, Sally (2016a). What is a (social) structural explanation? *Philosophical Studies*, *173*(1), 113-130.

Haslanger, Sally (2016b). Comments on Madva (2016). *Ergo* symposium.

He, J., Butler, B. & W.R. King. (2007). Team cognition: development and evolution in software project teams. *Journal of Management Information Systems 24*, 261-292.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews: Neuroscience 11*(7), 532-32.

Hoffman, D. & Tarzian, A. (2001). The girl who cried pain: a bias against women in the treatment of pain. *The Journal of Law, Medicine & Ethics, 29*, 13-27.

Hofmann, W., Gschwendner, T., & Schmitt, M. (2009). The road to the unconscious self not taken: discrepancies between self-and observer-inferences about implicit dispositions from nonverbal behavioural cues. *European Journal of Personality: Published for the European Association of Personality Psychology*, *23*(4), 343-366.

Holroyd, J. & Kelly, D. (2016). Implicit bias, character, and control. In A. Masala

and J. Webber (Eds.), *From Personality to Virtue: Essays on the Philosophy of Character*. Oxford University Press.

Howard, J.A. (1994). A social cognitive conception of social structure. *Social Psychology Quarterly*, 210-227.

Huebner, B. (2013). Socially embedded cognition. *Cognitive Systems Research*, *25*, 13-18.

Huebner, B. (2016a). Implicit bias, reinforcement learning, and scaffolded moral cognition. In Michael Brownstein and Jennifer Saul (Eds.), *Implicit Bias and Philosophy: Metaphysics and Epistemology* (Vol. 1, 47-79). Oxford University Press.

Huebner, B. (2016b). Transactive memory reconstructed: rethinking Wegner's research program. *The Southern Journal of Philosophy*, *54*(1), 48-69.

Huebner, B. (2016c). Socialized attention and situated agency. Keynote address, Minds Online Conference 2016.

Huet, E. (2016). Rise of the bias busters: how unconscious bias became Silicon Valley's newest target." *Forbes*. November 2, 2016. URL=<https://www.forbes.com/sites/ellenhuet/2015/11/02/rise-of-the-bias-busters-how-unconscious-bias-became-silicon-valleys-newest-target/#46994a0e19b5

Hyun, J. (2009). *Breaking the bamboo ceiling: career strategies for Asians*. HarperCollins.

Kawakami, K., Dovidio, J. & S. van Kamp (2007). The impact of counterstereotypic training and related correction processes on the application of stereotypes. *Group Processes & Intergroup Relations 10*(2), 138-156.

Lewis, K. (2004). Knowledge and performance in knowledge-worker teams: a longitudinal study of transactive memory systems. *Management Science 50*, 1519-1533.

Lewis, K. and Herndon, B. (2011). Transactive memory systems: current issues and

future research directions. *Organization Science 22*(5), 1254-1265.

List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics & Philosophy*, *18*(1), 89-110.

List, C., & Spiekermann, K. (2013). Methodological individualism and holism in political science: a reconciliation. *American Political Science Review*, *107*(4), 629-643.

Lowery, W. (2016). Aren't more white people than black people killed by the police? Yes, but no. *The Washington Post*. URL=<https://www.washingtonpost.com/news/post-nation/wp/2016/07/11/arent-more-white-people-than-black-people-killed-by-police-yes-but-no/>

Machery, E., Faucher, L. & D. Kelly. (2010). On the alleged inadequacies of psychological explanations of racism. *The Monist*, *93*(2), 228-254.

Madva, A. (2016a). A plea for anti-anti-individualism: how oversimple psychology misleads social policy. *Ergo 3*(27), 701-728.

Madva, A. (2016b). Replies to commentators. *Ergo* symposium.

Madva, A. (2017). Biased against de-biasing: on the role of (institutionally sponsored) self-transformation in the struggle against prejudice. *Ergo 4*(6), 145-179.

Mandelbaum, E. (2016). Attitude, inference, association: on the propositional structure of implicit bias. *Noûs 50*(3), 629-658.

Moreland, R.L., Argote, L. & R. Krishnan (1996). Socially shared sognition at work: rransactive memory and group performance. in J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 57-84). Thousand Oaks, CA, US: Sage Publications, Inc.

North Carolina Commission on Racial and Ethnic Disparities in the Criminal Justice System. (2015). Source: http://ncracialjustice.org/projects/implicit-bias-trainings/

Nosek, B., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101.

Ohlheiser, A. (2017). The woman behind 'Me Too' knew the power of the phrase when she created it – 10 years ago. *The Washington Post.* URL=<https://www.washingtonpost.com/news/the-intersect/wp/2017/10/19/the-woman-behind-me-too-knew-the-power-of-the-phrase-when-she-created-it-10-years-ago/?utm_term=.982cd5fff088>

Palermos, S.O. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research* 27, 25-41.

Pettit, P. (1993). *The common mind: An essay on psychology, society, and politics*. Oxford University Press.

Ren, Y. & Argote, L. (2011). Transactive memory systems 1985–2010: an integrative framework of key dimensions, antecedents, and consequences. *Academy of Management Annals* 5(1), 189-229.

Saul, J. (2016). Comments on Madva. *Ergo* symposium.

Skorburg, J.A. (2017). Lessons and new directions for extended cognition from social and personality psychology. *Philosophical Psychology*.

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences 9*(4), 465-481.

Trawalter, S., Hoffman, K. & A. Waytz (2012). Racial bias in perceptions of

others' pain. *PLoS ONE 11*(3): e48546.

The Washington Post (September 26, 2016). "Clinton on Implicit Bias in Policing." URL=< https://www.washingtonpost.com/video/politics/clinton-on-implicit-bias-in-policing/2016/09/26/46e1e88c-8441-11e6-b57d-dd49277af02f_video.html?utm_term=.5412a760d9c8dd49277af02f_video.html?utm_term=.3bfee8893a57?>

Wegner, D.M. (1987). Transactive memory: a contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). Springer: New York.

Wegner, D.M., Erber, R. & P. Raymond. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, *61*(6), 923-29.

Wilkinson, E. (2018). United we (don't) stand: how polarization manifests in sport. In J.A. Skorburg and W. Sinnott-Armstrong (Eds.), *Political Polarization and Morality*. The Kenan Institute for Ethics at Duke University.

Witt, C. (2011). *The Metaphysics of Gender.* Oxford University Press.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.