PSYCHOLOGICAL INQUIRY
2017, VOL. 28, NO. 4, 233–248
https://doi.org/10.1080/1047840X.2017.1335568

TARGET ARTICLE

The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice

B. Keith Payne[a], Heidi A. Vuletich[a], and Kristjen B. Lundberg[b]

[a]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [b]Department of Psychology, University of Richmond, Richmond, Virginia

ABSTRACT
As public awareness of implicit bias has grown in recent years, studies have raised important new questions about the nature of implicit bias effects. First, implicit biases are widespread and robust on average, yet are unstable across a few weeks. Second, young children display implicit biases indistinguishable from those of adults, which suggests to many that implicit biases are learned early. Yet, if implicit biases are unstable over weeks, how can they be stable for decades? Third, meta-analyses suggest that individual differences in implicit bias are associated weakly, although significantly, with individual differences in behavioral outcomes. Yet, studies of aggregate levels of implicit bias (i.e., countries, states, counties) are strongly associated with aggregate levels of disparities and discrimination. These puzzles are difficult to reconcile with traditional views, which treat implicit bias as an early-learned attitude that drives discrimination among individuals who are high in bias. We propose an alternative view of implicit bias, rooted in concept accessibility. Concept accessibility can, in principle, vary both chronically and situationally. The empirical evidence, however, suggests that most of the systematic variance in implicit bias is situational. Akin to the "wisdom of crowds" effect, implicit bias may emerge as the aggregate effect of individual fluctuations in concept accessibility that are ephemeral and context-dependent. This bias of crowds theory treats implicit bias tests as measures of situations more than persons. We show how the theory can resolve the puzzles posed and generate new insights into how and why implicit bias propagates inequalities.

KEYWORDS
Implicit bias; implicit prejudice; prejudice; systemic bias; racism

The notion of implicit bias can be puzzling. The idea is that people harbor mental associations based on race, gender, and other social categories that may lead to discrimination without intent, or possibly even awareness. Behavior as morally problematic as discrimination, without the kind of conscious intent that would bestow moral culpability, strikes many as perplexing in itself. In recent years, however, several research findings have raised new questions about the nature of implicit bias. In the pages that follow, we describe three empirical puzzles raised by implicit bias research. Then we suggest a new account of implicit bias that we believe can resolve these apparent contradictions and shift the way implicit bias is understood, both by theorists and by those aiming to reduce discrimination.

Three Puzzles

Puzzle 1: Large and Unstable

The first puzzle is that tests of implicit bias show robust average biases again and again. Yet, the temporal stability of these biases is so low that the same person tested 1 month apart is unlikely to show similar levels of bias. How can implicit biases be so reliable on average if the individuals composing those averages fluctuate so much?

Anecdotally, many educators know that classroom demonstrations of implicit tests, such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) and the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005), regularly reveal an evident bias, even in small groups. Replicable average biases have been documented more formally in studies with large Internet samples using the IAT (Nosek et al., 2007) and representative samples using the AMP (Payne et al., 2010), as well as in hundreds of studies that regularly report significant mean levels of bias using a variety of implicit tests (for reviews, see Fazio & Olson, 2003; Nosek, Greenwald, & Banaji, 2007). Although the reproducibility of many findings in psychology has been called into question, the finding that people, on average, display intergroup biases on implicit tests is not among them.

At the same time, the relatively low temporal stability of implicit tests has also been demonstrated in multiple longitudinal studies (e.g., Cooley & Payne, 2017; Cunningham, Preacher, & Banaji, 2001; Devine, Forscher, Austin, & Cox, 2012; for a review, see Gawronski, Morrison, Phills, & Galdi, 2017). In the most comprehensive study to date, implicit attitudes were measured (using the IAT and AMP) on topics of political attitudes, self-concept, and race (Gawronski et al., 2017). Average implicit bias effects were highly stable; means on both tests were within 2 percentage points across 2 months. Yet the implicit tests showed low test–retest correlations. Stability was particularly low for the implicit measures of race bias (IAT test–retest $r = .44$; AMP test–retest $r = .38$). Consistent with these estimates, Gawronski and colleagues (2017) meta-analyzed the published literature for studies reporting test–retest

CONTACT B. Keith Payne payne@unc.edu Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Campus Box #3270, Chapel Hill, NC 27599-3270, USA.

correlations for implicit measures and found an average stability of $r = .42$. This level of reliability suggests that less than 20% of the variability in implicit bias can be explained by an individual's level of implicit bias a few weeks earlier.

If individual levels of implicit bias are so capricious, why do samples converge so readily on similar mean effects? One possible explanation is that implicit tests are unreliable measures, so high error variance in the tests accounts for the other 80% of variability. However, the same studies demonstrating low test–retest correlations across measurement occasions have found reasonably high internal consistency for the IAT and AMP measures at each individual measurement occasion (e.g., Cooley & Payne, 2017; Gawronski et al., 2017; for reviews of implicit measure reliability, see Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Payne & Lundberg, 2014). This suggests that the tests are measuring something systematic within individuals at a single time point. Moreover, a study that used a latent variable approach to model and adjust for interitem consistency found that the stability was still low (Cunningham et al., 2001). Over 2 weeks, test–retest correlations were .36 for a response-window IAT, .46 for a response-time IAT, and .68 for a response-window priming task. Given that internal consistency cannot easily explain the low temporal stability of implicit bias measures, the most likely explanation is that the unreliability lies in the malleability of people's psychological biases rather than in the tests.

For most psychological measures, we would expect stable means because of stable individual scores. But the low stability estimates imply that the rank orders of participants (i.e., who is high on implicit prejudice and who is low) change dramatically from one measurement occasion to another, whereas the group mean scores somehow remain constant. Replicable implicit bias on average, paired with constantly shifting individual scores, presents an important puzzle to be explained.

Puzzle 2: Permanent Yet Unstable

A second puzzle follows from the first. Young children show levels of implicit bias similar to adults (e.g., Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2006; Dunham, Chen, & Banaji, 2013; Heiphetz, Spelke, & Banaji, 2013; for reviews, see Baron, 2015; Dunham, Baron, & Banaji, 2008). For example, in one study, implicit pro-White/anti-Black bias measured among White Americans using an IAT was virtually identical among 6 year olds, 10 year olds, and adults (Baron & Banaji, 2006; see also Dunham et al., 2013). Younger adults and older adults show similar levels of implicit bias once differences in executive control are accounted for (Stewart, von Hippel, & Radvansky, 2009). This developmental invariance is often interpreted as evidence that people learn implicit biases early and retain them for life (e.g., Dunham et al., 2013; see also Rudman, 2004b; Rudman, Phelan, & Heppen, 2007). Yet, if one's biases are not stable across a month, how can they be stable across a lifetime? Developmental invariance in constructs that are not temporally stable presents a second puzzle to be explained.

Puzzle 3: Places and People

The third puzzle is that individual differences in implicit bias predict behavior only weakly. Meta-analytic summaries of person-level associations between implicit measures and discriminatory behavior range from $r = .24$ (Greenwald, Poehlman, Uhlmann, & Banaji, 2009) to $r = .14$ (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013) for the IAT. Similarly, for sequential priming measures, the meta-analytic average association was $r = .28$ (Cameron, Brown-Iannuzzi, & Payne, 2012). Although the size of these associations depends on the measures used and the criteria for inclusion in the meta-analyses, the general conclusion seems clear: Individual difference measures of implicit bias are only modestly associated with individual differences in disparate treatment.

Some authors have interpreted these small effects of person-level biases to mean that implicit tests provide little meaningful information (Oswald et al., 2013; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Others have argued for the utility of implicit tests by observing that even small effects can have socially meaningful consequences (Greenwald, Banaji, & Nosek, 2015; Jost et al., 2009). However, there is another source of evidence that suggests implicit tests can predict important outcomes and with quite large effect sizes: Implicit bias is much more strongly associated with disparate outcomes when examined at aggregate levels such as nations, states, or metropolitan areas.

For example, in one study, countries with stronger average associations between males and science had greater gender-based achievement gaps in eighth-grade science and math scores (Nosek et al., 2009). Even after adjusting for several control variables, implicit bias remained strongly associated with achievement gaps in science ($\beta = 0.55$) and in math ($\beta = 0.67$). In another study, metropolitan regions in the United States with higher levels of implicit race bias were found to have greater racial disparities in shootings of citizens by police (Hehman, Flake, & Calanchini, in press). After controlling for demographic and geographical control variables, the association was $\beta = 0.39$. In France, city-level implicit bias toward Arab (vs. French) individuals was associated with lower participation rates in city marches aimed at national unity in the wake of an Islamist terrorist attack ($r = .37$; Zerhouni, Rougier, & Muller, 2016).

On the topic of health disparities (see Blair & Brondolo, 2017), one study found that death rates for Black residents were higher in U.S. counties in which Black residents had greater implicit biases against Whites ($\beta = 0.49$; Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016). Another found that counties with higher levels of implicit racial bias also had greater Black–White gaps in infant health outcomes, even when controlling for various demographic and geographical factors (Orchard & Price, 2017). Miller and colleagues (2016) found that in U.S. towns and town clusters with higher levels of implicit prejudice toward individuals with HIV, there were also higher levels of psychological distress among individuals with HIV in those communities (no standardized effect size estimates were reported).

These aggregate-level implicit bias effects appear to be strongly related to situational context as well. For example, U.S. states with higher proportions of Black residents in the population have higher average levels of implicit in-group preferences, among both White respondents ($r = .83$) and Black respondents ($r = .76$; Rae, Newheiser, & Olson, 2015). Moreover, the frequency of Internet searches for racial slurs is strongly associated with implicit in-group preferences for both White respondents ($r = .78$) and Black respondents ($r = .50$; Rae et al., 2015). These associations may reflect greater intergroup conflict in areas where Black and White residents are more likely to come into contact. Similarly, a study on nation-level weight bias found that those nations with a greater percentage of overweight individuals also had a greater average implicit preference for thin (vs. overweight) individuals ($\beta = 0.60$; Marini et al., 2013). (This trend was reversed at the individual level, with lower levels of implicit weight bias among those with a higher BMI, $\beta = -0.16$.) These substantial associations suggest that implicit bias may reveal more about geographical contexts than about individuals within them.

Together, these three puzzles—effects that are large yet unstable, biases that are paradoxically supposed to be permanent yet unstable, and biases that predict outcomes better for places than people—present important problems for the science of implicit bias to solve. These puzzles pose problems for theories that claim implicit biases arise from early experiences and/or are difficult to change because they are ingrained through years of experience (e.g., Petty, Tormala, Briñol, & Jarvis, 2006; Rudman, 2004b; Rydell & McConnell, 2006; Wilson, Lindsey, & Schooler, 2000). More generally, they pose problems for theories that treat implicit biases as instances of attitudes, knowledge, or belief, all of which imply some permanence (this includes virtually all psychological theories of implicit bias). We suggest a resolution to these problems, which begins with what we hope is an uncontroversial premise.

Concept Accessibility

Many theories assume that implicit attitudes reflect the accessibility of mental content (for a review, see Gawronski & Payne, 2010). Accessibility refers to the readiness with which information can be retrieved and used in cognitive processing (Fazio & Williams, 1986; Higgins, 1996; Srull & Wyer, 1979). Some theories assume that implicit evaluations reflect accessible associations, which are semantically connected mental content lacking logical relationships such as truth value (e.g., Gawronski & Bodenhausen, 2006). Others posit accessible propositions, which are statements about the world that include relationships and logical operations (e.g., De Houwer, 2014). Still others theorize a hybrid of both (e.g., Petty, Briñol, & DeMarree, 2007). Our model is agnostic about the format of representations that underpin implicit bias. We use the term *concept accessibility* to describe the likelihood that a thought, evaluation, stereotype, trait, or other piece of information will be retrieved for use. So, we propose that implicit bias reflects the accessibility of concepts linked to a social category. We use the term *accessible links* to describe the ease of accessing a concept, once a social category has been activated.

Different implicit tests may be designed to measure different types of accessible links. Suppose an AMP is designed in which Black and White faces are presented as primes, and participants evaluate target fractal images as pleasant or unpleasant. This task measures the accessibility of pleasant–unpleasant evaluations in response to race cues. Or, consider an IAT in which science-related words and arts-related words are to be sorted into joint categories with pictures of men and women. This task measures the relative accessibility of science versus arts concepts linked to gender categories.

Like other theories, we assume that, when we use one concept, other related concepts and information become more likely to be accessed for use and that this process happens spontaneously and involuntarily. When we read the words *bed, rest*, and *awake*, we make certain concepts more likely to be used to complete a word fragment such as sle___ (Roediger & McDermott, 1995). Likewise, when we read the words *Mohammed, mosque*, and *Islamic*, we make certain concepts more accessible to complete the thought, ter_____. This spontaneous activation of linked concepts can be modeled as spreading activation in a semantic network (e.g., Collins & Loftus, 1975) or as emergent patterns of activation in a localist connectionist network (Ferguson & Bargh, 2003; McClelland, 2000; E. R. Smith, 1996) or even as retrieval of compound cues shared between concepts in memory (McKoon & Ratcliff, 1992; Ratcliff & McKoon, 1988; for reviews, see Carlston, 2010; Payne & Cameron, 2013). The cognitive architecture underlying these processes is not our focus here. We only assume, like other theories of implicit evaluation, that using a concept spontaneously makes other related concepts more cognitively accessible.

Decades of research has pointed to a few primary factors that determine the accessibility of a concept. Concepts that are used more frequently and more recently become more accessible (Higgins, Rholes, & Jones, 1977; Rholes & Pryor, 1982; Srull & Wyer, 1979, 1980). Consistent with the assumptions of many theories, we assume that frequent and recent exposures to stereotypes and prejudices can forge connections between the mental representation of social groups and a variety of stereotypic traits and prejudiced evaluations. For example, the more frequently and recently one has encountered stereotypic portrayals of Black men as criminals, or women as fragile, or Muslims as terrorists, or White men as bigots, the more easily accessible these stereotypic attributes will be when one thinks about the social category.

The accessibility of links between a social category and a concept can vary both chronically (Higgins, King, & Marvin, 1982) and situationally (Higgins et al., 1977). For example, one person may have a link between the category of Black Americans and a negative evaluation that is more chronically accessible than that same link is for another person, which would presumably be captured by a higher implicit bias score. This notion of a chronically accessible link maps onto what most theorists mean when they discuss implicit attitudes and stereotypes. To the extent that implicit bias is driven by chronically accessible links, we would expect to see stable individual differences in implicit biases.

However, accessibility can also vary as a function of the situation. Numerous studies have documented that performance on implicit bias tests is malleable in response to various manipulations of the context. For example, implicit racial bias scores can be shifted by interacting with an African American experimenter, listening to rap music, or looking at a photo of Denzel Washington (Dasgupta & Greenwald, 2001; Lowery, Hardin, & Sinclair, 2001; Rudman & Lee, 2002). Other interventions that change implicit bias have focused on repeated pairings of category items with new associates (e.g., Karpinski & Hilton, 2001;

Olson & Fazio, 2006), the adoption of strategic responses such as implementation intentions (Mendoza, Gollwitzer, & Amodio, 2010; Stewart & Payne, 2008; Webb, Sheeran, & Pepper, 2012), practiced affirmations of counter-stereotypic pairings (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000), actual and imagined intergroup contact (e.g., Shook & Fazio, 2008; Turner, & Crisp, 2010), the experience and regulation of affective states (e.g., Dasgupta, DeSteno, Williams, & Hunsinger, 2009; Huntsinger & Sinclair, 2010), and taking the perspective of outgroup members (e.g., Todd, Bodenhausen, Richeson, & Galinsky, 2011; see also Lai et al., 2014). (For reviews, see Blair, 2002; Lai, Hoffman, & Nosek, 2013.)

Few, if any, of these interventions seem to have long-lasting effects, however (Lai et al., 2016). After a few hours to days, average levels of bias tend to snap stubbornly back to their baseline levels. One interpretation of this elasticity is that implicit bias scores reflect long-learned attitudes and beliefs. Although situational interventions can shift accessibility temporarily, people revert to their stable, chronically accessible attitudes. This explanation, however, assumes a stable attitude and is contradicted by the lack of stability in implicit biases. Findings that implicit biases are malleable, only to revert to their mean levels, are related to the puzzles presented earlier in this article. If there is little stability to individual biases, then why would scores return to a consistent mean level of bias?

So far, we have remained on common ground with most theories of implicit evaluation by assuming only that implicit biases are driven by accessible concepts linked to social categories and that accessibility can vary as a function of both situations and chronic individual differences. Here we depart this congenial territory to suggest an unconventional alternative interpretation.

The Bias of Crowds

Although concept accessibility can, in principle, vary both chronically and situationally, there is little empirical evidence for chronic accessibility that gives rise to stable individual differences in implicit intergroup bias. Instead, most of the systematic variance in implicit biases appears to operate at the level of situations. Situations could refer to an immediate social situation such as interacting with a Black experimenter during a lab study, or broader situations such as living in a particular place and time. It has long been recognized that a person's context and culture contribute to their implicit biases (e.g., Banaji, 2001). But this formulation still assumes that implicit bias resides as an attitude or belief with some degree of permanence in the minds of individuals and that the context is simply one kind of input that helps shape those attitudes. In light of the evidence just reviewed, we believe it is more accurate to consider implicit bias as a social phenomenon that passes through the minds of individuals but exists with greater stability in the situations they inhabit.

To summarize the view put forward here, although implicit bias can in principle exist as an attribute of persons or an attribute of situations, the empirical evidence is more consistent with the situational view. By switching the emphasis from a person-based analysis to a situation-based view, we arrive at a reinterpretation of the empirical data. This new interpretation suggests that measures of implicit bias are meaningful, valid, and reliable. Contrary to most assumptions, however, they are meaningful, valid, and reliable measures of situations rather than persons.

To clarify our view of implicit bias as a measure of situations, consider how a crowd of sports fans behaves when "the wave" erupts in one section of the stadium. The wave travels through the crowd within systematic parameters, with a stable velocity of about 20 seats per second (Farkas, Helbing, & Vicsek, 2002). Once the wave has begun in one section, we can predict with great accuracy whether, and when, a given section of fans will stand up. But if we tried to predict wave behavior by measuring attributes of individuals such as their personal propensity to stand or sit, we would likely be much less successful in our prediction. Far better to measure whether a wave was happening in that stadium at the time. When fans in one section all stand up at once and then sit back down, we have evidence that a wave has passed through that section. Moreover, if we followed those fans home, knowing that they stood and cheered at that moment would not predict whether they were likely to do so at the dinner table.

Viewing implicit bias as a social phenomenon that passes through individual minds, rather than residing in them, suggests another analogy that helps understand the empirical puzzles. Research on "the wisdom of crowds" (Surowiecki, 2004) finds that, for many kinds of questions, the collective judgment of a group tends to be closer to the true answer than any one individual's answer (Clemen, 1989; Galton, 1907; Lorge, Fox, Davitz, & Brenner, 1958; Mellers et al., 2014; Page, 2007). If you show a crowd a jar of jelly beans and ask how many there are, the average of the individual guesses is likely to be more accurate than even the most accurate individual (Surowiecki, 2004). Or, if you ask a sample of people the likelihood that some future event will occur, their aggregated estimates will tend to be more accurate than that of individuals working independently (Mellers et al., 2014). The "wisdom of crowds" effect has been used to improve prediction of political, military, and economic events by drawing on partial knowledge that is distributed across a population (Tetlock & Gardner, 2015).

Crowds are "wise" because each individual is likely to have partial true knowledge as well as erroneous biases that are largely random. When independent judgments are averaged, the random variations are aggregated away, leaving the true knowledge to emerge as the central tendency of the distribution. For any single person, that true knowledge might be fleeting and changeable. Before looking at an urn of marbles or considering the likelihood of a geopolitical event, a respondent may have never thought about the question before. And the next day, they might have forgotten, or changed their mind. But whatever partial information is available to the sample at the time of judgment is sufficient to create surprisingly accurate and stable estimates in the aggregate.

In studies of the wisdom of crowds, the questions of interest have an objectively correct answer. Implicit bias, in contrast, concerns largely subjective attributes such as evaluations and stereotypes of social groups. Nonetheless, the processes underlying the effects may have important similarities: To begin, we assume that all members of a culture have similar knowledge of the stereotypes and prejudices present in that culture (Devine, 1989; Katz & Braly, 1933). In making that same argument, Devine (1989) proposed that stereotypes are like language in that all members of a society share knowledge about those categories. Like Devine, we suggest that near-universal knowledge of stereotypes creates the potential for anyone to experience implicit bias. However, the language metaphor implies that an individual's implicit bias should be extraordinarily stable across time and contexts, because linguistic structures such as grammar and syntax do not change from one day to another. In contrast, we argue that mere knowledge (i.e., availability) of the stereotypic links is not sufficient for implicit bias to emerge. Instead, the accessibility of those concepts is critical, and accessibility fluctuates from one situation to the next.

Suppose that when a sample of research subjects completes an implicit test of racial evaluations, their score reflects the net accessibility of all attributes linked to the social categories of Black people and White people. Some elements of that concept accessibility will be widely shared among all residents of the same culture. Other elements will be specific to whatever each subject was doing and thinking in the moments before the test. Still other elements will be intermediate, reflecting partially shared concepts that may be highly accessible in some contexts and not in others. When aggregated across a sample of subjects, the average bias score will reflect the knowledge with the most widely shared accessibility. The idiosyncratic associates will be averaged away because they are randomly distributed across persons. In general, the most widely shared associates will tend to be those that are held in common due to stable contexts such as states or nations of residence. However, if the entire sample has been exposed to the same situational context by, say, watching a Denzel Washington movie, then the average level of bias should be shifted accordingly.

Most of us go to sleep most nights in the same town, state, and country where we woke that morning. But the social interactions, media exposures, moods, memories, and trains of thought that form our more immediate contexts shift from one situation to the next. Individual implicit bias scores will be influenced by all of these factors and more. But like the wisdom of crowds, aggregate responses should converge on the most widely shared influences. Implicit bias can therefore be thought of as the bias of crowds.

Scope and Limits of the Model

The arguments articulated here apply specifically to the concept of implicit biases toward social groups. They do not question the notion that people hold attitudes in general. Some scholars have argued that all attitudes are temporarily constructed evaluations rather than stable dispositions or knowledge structures (Schwarz, 2006, 2007). In Schwarz's view, the stability typically seen in attitude measures derives from the fact that people access the same "inputs" (i.e., information activated by the environment or retrieved from memory) each time the attitude is measured. Thus, even though an evaluation may be constructed on the spot, people tend to draw on similar information across time with which to construct it.

Schwarz's view of attitude construction is a theoretical perspective rather than an empirical claim. The debate between attitudes-as-dispositions and attitudes-as-constructions depends on the semantic question of what counts as "the attitude" and what counts as the "inputs." For example, Fazio (1995, 2009; see also Fazio, Chen, McDonel, & Sherman, 1982) defined attitudes as evaluative knowledge stored in memory. Schwarz (2007), in contrast, considered knowledge stored in memory and information that is chronically accessible to be stable aspects of the person, but they are inputs rather than attitudes. Any evidence used to demonstrate that attitudes are stable aspects of a person can also be interpreted as consistent with the constructionist perspective by changing where we draw the line between the unobservable entities of attitudes and inputs.

Schwarz (2007) argued that in this situation, it is more parsimonious to omit the concept of "attitude" as a theoretical construct and focus instead on the information and memory processes underlying evaluative judgments. Attitude theorists disagree that this is the more parsimonious option and argue that the concept of an attitude is an efficient summary of evaluative knowledge including affective, cognitive, and behavioral components (e.g., Fazio, 2009). Although these semantic debates are valuable for examining theoretical assumptions in the attitudes literature, they are not germane to the argument that we are advancing.

Our situationist view of implicit bias does not deny the view that attitudes exist. If people store evaluative knowledge in memory and that knowledge can be chronically accessible, then we would interpret that stability as evidence of dispositions that deserve to be called attitudes. Our model is instead a response to empirical findings that fit poorly with the traditional attitude view where implicit intergroup bias is concerned. Other kinds of implicit evaluations may operate much more as personal dispositions and, therefore, attitudes. For example, Gawronski and colleagues (2017) found much higher test–retest stability for implicit evaluations of political candidates than for racial groups. Other research has found that individual differences in implicit political attitudes are substantially associated with individual behaviors such as voting and policy preferences (Hawkins & Nosek, 2012; Lundberg & Payne, 2014). These findings suggest that some kinds of implicit evaluation reflect stable aspects of the person.

Why might implicit intergroup biases be more tied to situations and less tied to persons than some other kinds of implicit evaluations? We can speculate that the abstract nature of social categories makes intergroup biases especially unstable and malleable. Categories are, by definition, abstractions (Medin, 1989; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). A given category is linked to many different exemplars and attributes, but the category label summarizes some regularity about them. Because of this abstraction, evaluations of categories such as "men" and "women" or "Black people" and "White people" may be more context-dependent than evaluations of a single person or object. This may be why activating particular exemplars such as Denzel Washington or Martin Luther King can shift category evaluations. We have argued that people are conduits for implicit bias, which passes through them in the form of (largely temporary) concept accessibility. One implication of this argument is that implicit bias tests are more valid and reliable measures of situations than of persons. If our account is correct, then the three puzzles with which we began this article dissolve.

Puzzles Resolved

Findings of large average bias effects despite low temporal stability follow naturally from our account. We assume that individuals have a variety of accessible links to social categories, many of which are fleeting and changeable from one context to another, which makes individual scores variable and unreliable across time. The robust average effects arise because there is a high degree of prejudice in contemporary culture, communicated in countless ways, from biased depictions in media, to segregation in everyday interactions, to daily observations of which social groups tend to occupy high-status and low-status positions, and so on. Inequality in the culture exerts a constant influence that gives rise to large average effects, like a slow moving wave across generations.

These dynamics can explain why young children display the same biases as adults, even though individuals show little stability. When children and adults are sampled from the same times and places, they reflect the same stereotypes and biases that inhabit that context. Concretely, this means that they are cued by the same environments and therefore have the same concepts accessible, on average, at any given moment. This explanation contrasts with theories that interpret implicit bias as associations learned early, and therefore ingrained through years of learning (Petty et al., 2006; Rudman, 2004b; Wilson, Lindsey, & Schooler, 2000). These theories assume that information learned earlier is overlearned, and therefore is more likely to be automatically accessible. However, this assumption is inconsistent with the low temporal stability of implicit bias measures (Gawronski et al., 2017). Understanding implicit bias as a measure of situations reframes the finding that children and adults show similar biases. Rather than a finding of developmental invariance, this finding does not reflect development at all. Children and adults drawn from the same context simply display the same biases (Baron, 2015). If they were to change contexts, their implicit biases would presumably change with them. Like sports fans sitting together, children and adults express the same waves.

Our situational account also resolves the puzzle that implicit tests of bias are more predictive of outcomes at the aggregate level than the individual level. The level of bias varies in systematic ways across different counties, metropolitan areas, states, and nations, and implicit tests appear very effective at detecting these differences. The view of implicit bias as the bias of crowds helps integrate implicit bias research with research on structural and institutional racism and sexism, as we describe next.

Integrating Implicit Bias with Systemic Bias

Traditional psychological approaches to stereotyping and prejudice define prejudice as group-based evaluations, stereotypes as group-based beliefs, and discrimination as group-based disparate treatment (Allport, 1954). These evaluations, beliefs, and behaviors are considered to be rooted in the minds of individuals, encouraging studies of individual differences in attitudes, ideologies, and personality traits. Research in sociology has focused on the same topics but located racism and sexism in social systems rather than individual minds. Institutional racism, for example, describes norms and policies in institutions such as legal systems, educational systems, or mass media environments that favor powerful racial groups over less powerful groups. Systemic or structural racism is a related, but broader, concept focusing on the ways that racial hierarchies are built into history and culture in ways that pervade nearly all aspects of a society (Feagin, 2006; Feagin & Feagin, 1986). Scholarship on systemic prejudice has highlighted that in a racist or sexist system, women and minorities will be systematically disadvantaged regardless of the attitudes of individuals.

Hurricane Katrina in 2005 provided a vivid example of how systemic bias can operate. Although the hurricane itself was a random misfortune, the flood that followed selectively damaged lower elevation areas which were disproportionately inhabited by poor and Black residents of New Orleans. Thus, historical inequalities in property ownership and residential segregation had the effect of perpetuating disadvantage along income and racial lines. The evacuation plan was based on interstate highways, which assumed that residents had cars, further putting residents without cars—primarily poor and Black residents—at greater risk (Powell, 2008). Systemic bias in the absence of intentional discrimination leads to what sociologists have called "racism without racists" (Bonilla-Silva, 2003). This idea is strikingly resonant with psychology's notion of implicit bias, although operating at different levels of analysis.

There is little consensus on how to measure systemic bias, and authors have used a variety of social indicators to capture it. Examples of such indicators include measures of residential segregation, neighborhood pollution, race- and gender-based disparities in incomes and wealth, educational achievement gaps, differential incarceration rates, content analyses of media portrayals of social groups, and many other metrics (Cole & Farrell, 2006; Fryberg, Markus, Oyserman, & Stone, 2008; Gee & Ford, 2011; LaVeist, 1989; Lukachko, Hatzenbuehler, & Keyes, 2014; Mandel & Semyonov, 2005; C. D. Smith, 2009; van Dijk, 1989; Williams & Collins, 2001). From an individual-based perspective, these disparities reflect the cumulative outcomes of prejudice and discrimination among myriad individuals over time. But from a systems-based perspective, these disparities constitute the racism and sexism itself.

Scholarship on implicit bias takes for granted that systemic bias exists, and it is typically assumed to contribute to implicit bias (Greenwald & Krieger, 2006; Jost & Banaji, 1994; Rudman, 2004a). Where else but a biased culture would we learn biased associations? But in psychology research on implicit bias, the focus has remained squarely on individuals and, in particular, individuals with high levels of implicit bias. This may be because notions of systemic prejudice, distributed throughout culture and history and institutions, seems too vague to be quantified and rigorously tested. To some scholars of systemic racism, in contrast, the focus on the individual is considered reductionist because it neglects the historical and cultural roots of prejudice (Freeman, 1978; Wellman, 2007). Although a number of theorists (Berard, 2008; Richeson & Sommers, 2016) have called for better integration between these different levels of analysis, scholarship on implicit bias and systemic prejudice have proceeded largely in parallel with little cross-talk.

Our bias of crowds model serves to integrate the two approaches in two ways. First, we assume that the average level of bias in a situation or environment reflects the net frequency and recency of stereotypical links made accessible in that context. If so, then implicit bias is a psychological marker of systemic prejudice in the environment.

Suppose that Town A has relatively high levels of systemic racism. Housing patterns and schools are highly segregated, and they are correlated with large disparities in incomes. Because of the economic inequalities and segregation, crime is concentrated in the mostly poor and mostly minority areas. A person walking through Town A will see mostly White teachers, doctors, and bankers, and mostly non-White service workers operating cash registers and emptying the trash. When police pull over motorists, or criminal suspects are described in the local news, they are disproportionately Black or Hispanic. Town B, in contrast, has low levels of systemic racism. Residents know all the same stereotypes as everyone else. But strolling through the town, residents are unlikely to see those stereotypes confirmed by inequalities in living conditions and social roles on a daily basis. Because of the difference in daily reminders of inequality, the average accessibility of stereotypical links will be different in the two towns. Implicit bias will be higher in Town A than Town B.

In addition to functioning as a marker of systemic prejudice, implicit bias can also serve as a mechanism that translates systemic prejudice into individual discrimination. Much research has shown that when stereotypic thoughts or feelings are activated at the time of a judgment, the judgment tends to be biased toward that stereotype (Banaji & Greenwald, 1995; Banaji, Hardin, & Rothman, 1993; Bargh, Chen, & Burrows, 1996; Bless, Schwarz, Bodenhausen, & Thiel, 2001; Gilbert & Hixon, 1991; Higgins et al., 1977; Lepore & Brown, 2002; Srull & Wyer, 1979). The evidence just reviewed on geographical variability suggests that stereotypical thoughts and feelings will tend to be more accessible in regions with higher levels of structural prejudice. Individuals in those contexts should therefore be more likely to make biased decisions and judgments in large and small ways on a daily basis. Although this mechanism operates in the minds and decisions of individuals, it is not limited to individuals with chronically high levels of implicit bias. Rather, individual minds reflect the inequalities they see around them on average, in real time. Individual implicit bias scores may fluctuate based on what they have witnessed and thought about in the hours and minutes before completing the implicit test. But the average score in Town A will still be higher than the average score in Town B. This difference in accessible concepts predisposes more people in Town A to discriminate at any given moment than in Town B.

Implications for Debates in Implicit Bias Research

A few questions have been debated in implicit bias research since it began. Here we consider some of the central debates from the bias of crowds perspective. These questions concern whether implicit bias reflects personal attitudes or extrapersonal associations, whether people are consciously aware of their implicit biases, and how best to reduce the discrimination that results from implicit bias.

Personal Attitudes or Extrapersonal Associations?

One critical question has concerned whether measures of implicit bias reflect personal attitudes or "extrapersonal associations" such as cultural stereotypes. In an early study, Karpinski and Hilton (2001) found that participants showed an implicit preference on an IAT for apples over candy bars, but when asked to choose they overwhelmingly chose candy bars. There was no association between individual differences in IAT scores and their behavioral choice. The authors argued that the IAT may reflect associations in the culture that value healthy foods over candy, but those associations have little to do with participants' personal attitudes. In another study, Han, Olson, and Fazio (2006) found that the uninformed preferences for Pokémon characters expressed by children influenced IAT evaluations of those characters. They interpreted this finding as evidence that the IAT's attitude score is contaminated by extrapersonal associations.

The question of whether implicit tests reveal personal attitudes versus extrapersonal associations has been debated mainly at the level of particular measurement tasks. To improve the IAT's ability to measure personal attitudes, Karpinski and Steinman (2006) modified the IAT to use a single category, and Olson and Fazio (2004) modified the IAT to require more personalized categorizations of the stimuli (e.g., "I like/I don't like" as opposed to "pleasant/unpleasant"). These modified tasks showed better individual difference associations with outcome variables and less susceptibility to environmental influences than the standard IAT, but the implications of these measures for the present model are unknown. First, these studies focused primarily on concrete attitude objects such as apples, candy bars, flowers, insects, and cartoon characters. Intergroup biases based on race and gender categories are more abstract, as discussed earlier, and may be more easily influenced by cultural associations. Second, little is known about whether modified versions of the IAT or other implicit measures show stability over time, as would be expected for measures of individual attitudes. Future research may reveal that modifications to implicit tests can produce measurements that capture individual-level prejudices and stereotypes, but the existing literature has not provided such evidence.

In an influential critique of implicit bias research, Arkes and Tetlock (2004) argued that if implicit measures capture stereotypic associations that individuals reject but are aware that other people hold, then implicit measures are not capturing prejudice in a meaningful sense. They illustrated their point using an anecdote from civil rights leader Jesse Jackson, who described a feeling of relief when he heard footsteps behind him and the person turned out to be White. Their argument was that if someone who has devoted his life to fighting racism can show the kinds of cognitive and emotional responses that are characteristic of implicit bias, then implicit bias must not really be prejudice. Demonstrating prejudice, according to Arkes and Tetlock, requires a personal attitude that is "grounded in hostility" and "unwarranted as well as resistant to change" and "that the affective negativity tapped by implicit associative measures does not merely reflect culturally shared associations that might arise in any society with widespread inequality" (p. 258).

Banaji and colleagues (2004) replied that drawing a bright line between cultural associations and personal attitudes is artificial because people learn their attitudes and beliefs from their culture (see also Banaji, 2001). Moreover, they argued that the kind of overt hostility referenced by Arkes and Tetlock describes explicit prejudice rather than implicit prejudice, which consists of mental associations that need not be endorsed to influence thought and behavior. Consistent with the argument made here, Banaji and colleagues considered implicit prejudice to be associations that come to mind and can lead to discriminatory outcomes, regardless of intent.

This debate dissolves when viewed from the bias of crowds perspective. Like Banaji and colleagues, we view implicit bias as being rooted in cultural associations and not dependent upon intent. However, we argue that there is a meaningful distinction between cultural (or environmental or situational associations) and personal implicit attitudes. Environmental differences in structural and institutional bias give rise to patterns of construct accessibility that reflect cultural stereotypes. These accessible concepts can vary both chronically (which corresponds with Banaji and colleagues' notion of implicit prejudice) and situationally (which corresponds to our metaphor of a wave passing through a crowd). If biases learned through one's immediate context create a lasting change that stays with a person as he or she changes situations over time, then we would label this chronic accessibility as a personal implicit bias. As just reviewed, both chronic and temporary accessibility are possible, but the empirical evidence on temporal stability and predictive validity appears stronger for the situational view.

From the bias of crowds perspective, describing implicit bias as "culturally shared associations that might arise in any society with widespread inequality," as Arkes and Tetlock (2004, p. 258) did, is not a criticism of the concept, because accessible stereotypes can increase the likelihood of acting in biased ways (Bodenhausen, 1988; Srull & Wyer, 1979). Widespread inequalities is another way to describe systemic prejudice, and associations passing through the mind of individuals are one way that systemic prejudice manifests at the psychological level. From our perspective there is nothing contradictory about Jesse Jackson's anecdote. Any given person—even a Black civil rights leader—steeped in a culture with high levels of systemic bias will have a high likelihood of stereotype-consistent thoughts and feelings passing through his mind.

Is Implicit Bias Unconscious?

Implicit bias is often assumed to refer to attitudes that people are unaware that they hold. In an influential early theoretical article, Greenwald and Banaji (1995) defined implicit bias as an "introspectively unidentified (or inaccurately identified) trace of past experience" that mediates biased responses (p. 5). A number of authors have explicitly argued that people cannot introspect upon implicit attitudes (e.g., Devos, 2008; Kassin, Fein, & Markus, 2011; Kihlstrom, 2004; McConnell, Dunn, Austin, & Rawn, 2011; Spalding & Hardin, 1999). Many others have implied the same conclusion by using the terms *implicit* and *unconscious* interchangeably (e.g., Bosson, Swann Jr, & Pennebaker, 2000; Cunningham, Nezlek, & Banaji, 2004; Jost, Pelham, & Carvallo, 2002; Phelps et al., 2000; Quillian, 2008; Rudman, Greenwald, Mellott, & Schwartz, 1999). These claims have typically been based on low correlations between implicit and explicit measures of the same topic. Implicit–explicit correlations are especially low for racial bias (e.g., Cameron, Brown-Iannuzzi, & Payne, 2012; Hofmann et al., 2005; Nosek, 2005, 2007; Nosek & Hansen, 2008; Nosek & Smyth, 2007). These low correlations with self-report have bolstered the common assumption that if concepts are assessed by an implicit test, then the concepts must be unconscious.

Correlations between implicit and explicit measures could be low for many reasons other than a lack of conscious awareness, however, and there is little evidence that implicit attitudes are unconscious (Gawronski, Hofmann, & Wilbur, 2006). More recent research asked subjects to predict their performance on several IATs by judging how difficult it would be to sort various combinations of stimuli and categories together. Predictions were highly accurate, including on tests of implicit racial bias (Hahn, Judd, Hirsh, & Blair, 2014). This finding suggests that, at least under certain conditions, people can introspect and accurately report about their implicit biases. Note, however, that participants in this study were asked to predict the difficulty of the task rather than express their attitudes or beliefs. What people are asked to introspect about may be critically important for whether they accurately report on implicit biases.

In another series of studies, participants were asked to self-report on their "considered opinions" and their "gut reactions" toward gay and straight people (Ranganath, Smith, & Nosek, 2008). Reports of "gut reactions" were substantially correlated with performance on three implicit bias tests, but reports of "considered opinions" were not. Participants were also asked to introspect about their attitudes toward gay people and consider how their feelings unfold over time. Then they rated how they would feel at several time points, ranging from their "instant reaction" to when "given time to think fully about my feelings." Consistent with performance on speeded response tasks, participants reported that they would be more biased against gay people in their instant reactions than when given time to think carefully (Ranganath et al., 2008). These studies suggest, again, that the accuracy of self-reports depends on exactly what mental content people are directed to report about.

From the bias of crowds perspective, the malleability in self-report accuracy stems from the fact that for any individual, the accessibility of various stereotypic concepts is continuously fluctuating. These momentary fluctuations drive performance on implicit tests at that moment. When introspecting, people can report about the thoughts and feelings that are passing through their minds. They are most likely to do so if they are asked appropriate questions, such as to rate their "gut reactions" or their "instant reactions." However, the concepts that are active at that moment may have little to do with what subjects consider to be their attitudes and beliefs. For this reason, asking them to rate their "considered opinions" or their beliefs or values would lead them to report about a set of information very different from the concepts that happen to be most accessible at the moment.

When you are seated in a sports arena and the wave passes through the stands, it is clear what you are supposed to do when the wave reaches your seat. You know what to do, but what does the wave *mean*? And did you stand up *intentionally*, or did it just happen? These are curious questions because, as an individual, the wave may not have any clear meaning. After standing up as part of the wave, one could just as reasonably say that he or she intentionally chose to stand up (I wanted to be a part of it) as to say that he or she felt compelled to stand up by an external force (the wave made me do it). When social phenomena pass through our minds, they are often ambiguous stimuli that have to be interpreted.

Implicit biases may work the same way. Imagine that you are in an airport when you see a young man in traditional Muslim dress and thoughts about terrorism briefly cross your mind. If an experimenter stopped you at that moment and asked you to complete an implicit test of anti-Muslim bias, you would likely score high. But if the experimenter gave you a questionnaire that asked you to report on your "deeply held beliefs" about Muslims, you might report different beliefs. The research just reviewed suggests that if you were asked more specific questions, such as to rate your "gut reactions" or to describe the thoughts that "passed through your mind," you might admit more biased thoughts. But there is a wide range of other ways to interpret those spontaneous thoughts as feelings as well. Research suggests that accessible thoughts and feelings that drive implicit biases are ambiguous stimuli that must themselves be interpreted in order to be reported.

In one study, subjects completed an AMP to measure implicit biases toward gay people, then one group was led to interpret any feelings they experienced as their intentional thoughts and feelings, whereas another group was led to interpret them as unintentional reactions (Cooley, Payne, & Phillips, 2014). Finally, all participants completed an explicit measure of antigay attitudes. Participants scoring high on implicit bias reported greater explicit prejudice when they were led to interpret their implicit bias as intentional. Another study used the same design to manipulate perceptions of ownership. One group was led to believe that whatever thoughts and feelings they had toward gay people enduring the implicit test reflected their "personal attitudes," whereas the other group was led to believe that they reflected "societal stereotypes." Participants with high implicit bias reported much more explicit prejudice when they were led to interpret accessible thoughts and feelings as personal attitudes (Cooley, Payne, Loersch, & Lei, 2015).

A similar design was used to manipulate the emotional experiences arising from implicit bias (Lee, Lindquist, & Payne, in press). After measuring implicit bias toward Black Americans, participants were led to interpret any negative affect they experienced as either fear of Black people or sympathy for Black people. Following the manipulation, participants led to interpret their affective responses as fear reported greater explicit fear of Black people, displayed greater skin conductance in response to pictures of Black men, and perceived greater anger on the faces of Black men.

These studies suggest that the thoughts and feelings revealed by implicit bias tests can be interpreted in a variety of ways, with important consequences for downstream behavior. From a bias of crowds perspective, thoughts and feelings that spontaneously pass through our minds are easily interpreted in a variety of ways because they have no fixed meaning. They are simply concepts that came to mind, often in response to environmental cues.

In sum, the bias of crowds perspective suggests that people can introspect and report about currently accessible concepts. In that sense, subjects are aware of their implicit biases. However, they may not experience those accessible concepts as arising from stable attitudes and beliefs. As a result, researchers often mistakenly conclude that subjects are unconscious of their implicit biases because they ask subjects to report about their attitudes, beliefs, or considered opinions when subjects experienced only a stray thought or feeling. Better understanding how people interpret accessible thoughts and feelings would shed further light on the relationship between implicit bias and conscious experience.

How Can Discrimination Best Be Reduced?

A situationist view of implicit bias has clear implications for efforts to reduce the discrimination and inequality that result from it. Implicit bias training has become a widespread tactic to fight discrimination. The goal of such training is to change people's implicit attitudes. Interventions aimed at changing implicit bias have shown short-term efficacy. Yet, average levels of implicit bias return stubbornly to baseline after a delay of days or weeks (Chapman et al., in press; Devine et al., 2012; Lai et al., 2016). From the bias of crowds perspective, interventions that attempt to change implicit attitudes are likely to have limited success because most of the force of implicit bias comes from situations and environments rather than individual attitudes. Individuals functioning within biased situations will tend to return to the average level of bias in the situation.

A second goal of implicit bias training is to raise awareness of one's own implicit biases in order to exert control over them. Based on the research just described, raising awareness could have either helpful or counterproductive effects, depending on how participants make sense of their implicit biases. If participants come to view their implicit biases as reflecting their true personal beliefs or their intentional responses, they are more likely to endorse them explicitly and bolster the sentiments that came spontaneously to mind (Cooley et al., 2015). If they come to view them instead as cultural stereotypes or unintended responses, they are more likely to explicitly reject them (Monteith, Voils, & Ashburn-Nardo, 2001). Thus, the effectiveness of consciousness raising may depend on the phenomenal quality of conscious experiences.

Guided by the view that implicit bias reflects early learning of cultural biases that lead to stable individual differences in implicit attitudes, some authors have called for using implicit tests to screen out highly biased individuals from roles in which discrimination would be likely or would be especially harmful (e.g., Bennett, 2010). Scholars who study implicit bias have generally opposed the use of implicit measures for screening purposes on the grounds that the associations between individual difference measures and individual discriminatory outcomes are too small to make decisions about particular individuals (Schnabel, Asendorpf, & Greenwald, 2008). We certainly agree, but the recommendation following from a situationist view goes further. If implicit bias arises from situations, then it may be useful to "screen" situations and contexts rather than people. More research is clearly needed before such practices are implemented. The available data, however, suggest the hypothesis that average levels of implicit bias at the level of organizations, such as firms, universities, department, or offices, may predict the likelihood of disparate outcomes better than individual scores. Of course, what it means to "screen" or make "selection" decisions is different at the organizational level than the individual level. If high levels of implicit bias are detected in an organization, it may be a sign that the organization should take steps to remediate the impact of implicit bias on decisions and outcomes.

If implicit biases arise from systemically biased situations, then the most effective route to reducing discrimination is to structure situations better. Although a full discussion of remediation strategies is beyond the scope of this article, a few general recommendations for structuring situations can be gleaned from the literature and our reinterpretation of it. Each is useful as a general practice wherever the potential for bias exists but would be particularly valuable in contexts where average levels of implicit bias are higher.

First, the general findings that most people show substantial levels of implicit bias on average suggest that decision processes should use blind review practices as much as possible. Hiding the social categories of people being evaluated is a simple way to reduce the potential for all kinds of biases. Second, people tend to make biased decisions and then confuse their reasons for their rationalizations (Norton, Vandello, & Darley, 2004). Using a rubric for important decisions in which the important criteria are spelled out beforehand is one practical way to prevent post hoc rationalizations (Uhlmann & Cohen, 2005). Stereotypes are most likely to bias decisions under conditions of ambiguity, when it is not clear what the right answer is or, for example, who the best job candidate is. A rubric with clear a priori criteria helps prevent exploiting such ambiguity under vague categories such as "poor fit."

Implicit biases are most likely to affect judgments and decisions when the opportunity for more thoughtfully controlled processing is limited (Payne, 2005, 2006; Sherman et al., 2008). Structuring decision settings in a way that minimizes rushing, fatigue, and distraction can reduce the potential for bias. For example, allocating ample time to review large sets of résumés can prevent rushed responding. Making important decisions during a time of day when decision makers are most alert can minimize fatigue. Minimizing workplace distractions such as multitasking and noisy environments can help decision makers devote their attention to more intentionally controlled decision processes.

Finally, the observation that implicit bias tends to be present on average means that decision makers should recognize that the potential for biased decision making is the normal default condition in most situations. Recognizing that *bias is the baseline* means that colorblind strategies focused just on avoiding discrimination are unlikely to be sufficient to reduce disparities. From the bias of crowds perspective, the reason that implicit bias is widespread in general is that the environment has a relatively constant level of disparities and systemic inequalities that repeatedly raise the accessibility of stereotypic concepts. Thus, most environments are not neutral; they exert a constant level of pressure in the direction of stereotype-consistent bias. In contexts where the average level of bias is substantial, reducing disparities may require going beyond passive colorblind policies. Affirmative strategies for increasing inclusivity, diversity, and the visible presence of women and minority group members in positions of authority may be necessary to offset the constant "background radiation" of systemic bias that gives rise to widespread stereotype accessibility. The solutions proposed here are not unique to the bias of crowds perspective, but removing the focus from biased persons to biased situations highlights the relevance of policy-based interventions that influence social contexts, structures, and processes as opposed to individual attitudes. This emphasis is in keeping with social psychology's traditional focus on the power of the situation.

Novel Predictions and Implications

We began by showing how the bias of crowds model can resolve three puzzles in the existing literature. We end by describing novel predictions and implications of the model that can be tested in new research.

One implication concerns where researchers should look for evidence about whether, and how, implicit bias influences behavior. As just reviewed, most studies have taken an individual difference approach and found modest associations. The bias of crowds model suggests that manipulations or measures of situations and social contexts that influence implicit bias scores should also influence behavior. We have argued that implicit bias is an instance of concept accessibility. Several well-developed theoretical models exist for understanding the effects of activated information on judgments and behavior. These include models of concept accessibility in impression formation (e.g., Higgins et al., 1977) and behavior (e.g., Loersch & Payne, 2012), as well as models of how feelings guide judgments (Bless et al., 1996; Schwarz & Clore, 1983) and how people correct (or fail to correct) for activated information (Wegener & Petty, 1997). These models differ in their specifics, but they converge in observing that activated information is most likely to bias judgments and behavior (a) when the judgment is ambiguous, (b) when the source of activated information is either unknown or not considered an inappropriate source of bias, and (c) when people are either unmotivated or unable to correct their responses for the potentially biasing influence. Viewing implicit bias through a situational lens provides a reminder that the field already knows a great deal about how activated biases affect human behavior.

The important distinction between aggregate levels of implicit bias and individual differences means that multilevel data and multilevel analytical models are critical for evaluating the model. Estimating the effects of group-level biases on group-level behavior will require relatively large-scale multilevel data collection efforts. Data from Project Implicit have been invaluable in discovering the aggregate effects reviewed here because they provide multilevel comparisons. Additional large-scale collaborative data collection efforts should be developed to test the multilevel effects of situations on implicit bias and behavioral effects that may flow from it.

In addition to geographical variability, the role of situations and persons can be estimated using intensive longitudinal data in which subjects complete implicit measures many times over days, weeks, or months. The bias of crowds model suggests that scores will be highly variable from one measurement occasion to the next, but that other variables, such as behavioral outcomes, should also fluctuate in a similar pattern. This approach may help isolate the systematic situational variability driving implicit tests that is often mistaken for error variance from an individual difference perspective.

A further, surprising, implication follows from the idea of implicit bias as a measure of situations. The average level of bias in a situation or context should be predictive of what outcomes are likely to come to pass, even if the individuals tested are not involved. Consider the finding described earlier, that metropolitan areas with greater average implicit race bias have greater racial disparities in police shootings. The vast majority of respondents whose implicit bias was measured in this study were not police officers. Yet Black residents living in an area where a representative sample of subjects show a high level of bias are more likely to be shot—by other people.

This is deeply strange from the view of implicit bias as an individual attitude. But it follows naturally if implicit bias reflects bias in the environment. We assume that the average level of implicit bias in a region reflects the average probability of having biased accessible links activated for any given person and any given moment. If the same environmental influences make racially biased links accessible for most people in that context, then the people for whom implicit bias is measured and the people making discriminatory decisions do not need to be the same people. As with a wave passing through a crowd, we don't need to measure the behavior of every person to know what trends their behavior will reveal.

Another novel prediction concerns the conditions under which implicit biases will appear to be conscious or unconscious. The bias of crowds view suggests that the momentarily accessible thoughts and feelings cued by social categories should be consciously reportable, if questions are posed appropriately. However, the model also suggests that because those accessible concepts are ambiguous stimuli, in many cases they will be interpreted by subjects in ways that differ from what researchers think they mean. Such misunderstandings between subjects and researchers will result in mismatches between what subjects report and what researchers are trying to measure, which will appear as "unconscious bias."

Suppose, for example, that a research subject is aware that she has stereotypical thoughts passing through her mind but does not think that means that she dislikes the group in question. The researcher, meanwhile, thinks that the presence of stereotypical thoughts does indicate prejudice. If that subject displays bias on an implicit test but reports low levels of prejudice on an explicit questionnaire, the stage is set for "introspectively unidentified (or inaccurately identified) traces of past experience" that could constitute unconsciousness (Greenwald & Banaji, 1995). However, we are in no position to know whether the inaccurate identification is on the part of the subject or the researcher. Distilling the "real meaning" of concept accessibility requires an act of interpretation—by both the subject and the researcher—and sometimes they will disagree.

Progress toward resolving these confusions over conscious awareness has been made by asking more focused questions to probe subjective awareness (Ranganath et al., 2008), providing subjects with more meaningful metrics on which to express their subjective responses (Hahn & Gawronski, 2014), and attending to structural fit between implicit and explicit measures to ensure that they are assessing the same stimuli in comparable ways (Payne, Burkley, & Stokes, 2008). Future research should attend to qualitatively different ways that subjects might experience the accessible concepts that constitute implicit bias, including a variety of emotions, thoughts, and other mental states (Cooley et al., 2015; Lee, Lindquist, & Payne, in press). The bias of crowds model predicts that if questions about subjective experiences are posed in ways that allow subjects to express how they interpret their own (often ambiguous and momentary) experiences of concept accessibility, their introspections will be highly accurate. But to the extent that subjects interpret their experiences with concept accessibility differently than the questions posed by researchers, implicit biases will appear to be introspectively unavailable.

Table 1. Bivariate correlation matrix of implicit race bias scores across U.S. states, from 2007 to 2016.

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2007	1									
2008	.679	1								
2009	.792	.690	1							
2010	.764	.761	.804	1						
2011	.770	.659	.794	.875	1					
2012	.728	.688	.738	.794	.859	1				
2013	.796	.702	.832	.809	.858	.760	1			
2014	.787	.658	.813	.762	.769	.754	.823	1		
2015	.626	.607	.727	.630	.670	.696	.755	.604	1	
2016	.694	.664	.780	.601	.640	.675	.718	.688	.782	1

Note. $n = 50$ states. All coefficients are significant, $p < .01$.

A final new prediction we discuss is that implicit bias measures should have a high degree of temporal stability, so long as we examine the stability of situations or contexts rather than persons. We are not aware of any research that has examined the stability of implicit bias at the level of situations or contexts. So as a first test, we analyzed state-level data from Project Implicit on average levels of implicit race bias over the past 10 years.

As predicted, we found very high levels of stability in the average level of bias across states (see Table 1). From year to year, the average test–retest correlation was $r = .76$. Across the full decade, the correlation was $r = .69$ (see Figure 1).

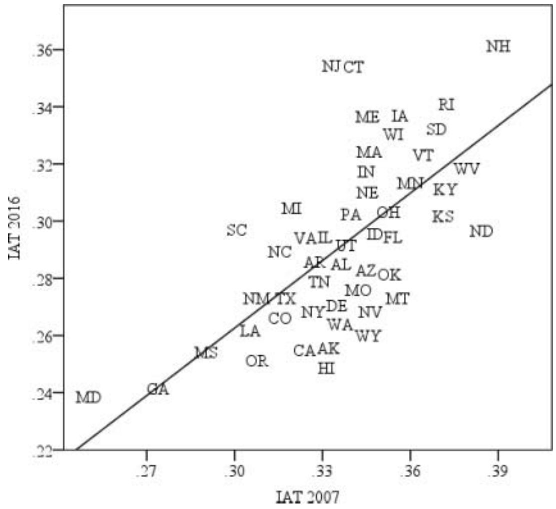

Figure 1. Average race Implicit Association Test (IAT) scores across U.S. states in 2007 and 2016. *Note.* Higher scores indicate a greater bias favoring Whites over Blacks. Units are *d* scores.

Individual levels of implicit bias are fickle across a month, but the states that have the highest and lowest levels of implicit bias tend to stay that way for years. This analysis has large numbers of subjects, ranging from 115 to 46,322 across states. One might argue that the superior stability found at the state level, as compared to individual scores, is an artifact of such large samples. Large samples within states, however, only increase the precision of the estimated means for each state. The slope of the line relating average implicit bias in 2007 to average implicit bias is 2016 is not inflated by large samples.

This high degree of stability provides important insight on why average implicit bias scores tend to return to a non-zero baseline again and again. Consistent with research on systemic bias, particular geographical contexts appear to have structures that systematically exert a pressure on average implicit bias scores. The level of pressure exerted differs in ways that are highly stable. The stability that researchers have been searching for in individual attitudes turns out to be found instead in contexts.

Conclusion

We have outlined a new perspective on the nature of implicit bias that has its roots in the very earliest scholarship on implicit bias. Researchers have always assumed that the widespread implicit biases documented in hundreds of studies are rooted in societal inequalities and the culturally shared stereotypes those inequalities breed. However, the person-centric assumptions dominant in this literature have led to many puzzles because the pervasiveness of implicit bias at the aggregate level has not been matched by high stability or predictive validity at the individual level. By returning to original ideas about the roots of implicit bias in systemically biased social structures, we offered a plausible solution to these puzzles. The bias of crowds model is consistent with widely shared assumptions about the psychological mechanisms underpinning implicit bias. The most novel aspect of the model is, in fact, its most orthodox: Understanding unintended discrimination requires appreciating the power of the situation.

References

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test.” *Psychological Inquiry*, *15*(4), 257–278. doi:10.1207/s15327965pli1504_01

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger III, J. S. Nairne, I. E. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, *68*(2), 181–198. doi:10.1037/0022-3514.68.2.181

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*(2), 272–281. doi:10.1037/0022-3514.65.2.272

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry*, *15*(4), 279–310. doi:10.1207/s15327965pli1504_02

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244. doi:10.1037/0022-3514.71.2.230

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives*, *9*, 50–54. doi:10.1111/cdep.12105

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58. doi:10.1111/j.1467-9280.2005.01664.x

Bennett, M. W. (2010). Unraveling the Gordian knot of implicit bias in jury selection: The problems of judge-dominated voir dire, the failed promise of Batson, and proposed solutions. *Harvard Law & Policy Review*, *4*, 149–171. Retrieved from http://harvardlpr.com/wp-content/uploads/2013/05/4.1_8_Bennett.pdf

Berard, T. J. (2008). The neglected social psychology of institutional racism. *Sociology Compass*, *2*, 734–764. doi:10.1111/j.1751-9020.2007.00089.x

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261. doi:10.1207/S15327957PSPR0603_8

Blair, I. V., & Brondolo, E. (2017). Moving beyond the individual: Community-level prejudice and health. *Social Science & Medicine*. Advance online publication. doi:10.1016/j.socscimed.2017.04.041

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology*, *71*(4), 665–679.

Bless, H., Schwarz, N., Bodenhausen, G. V, & Thiel, L. (2001). Personalized versus generalized benefits of stereotype disconfirmation: Trade-offs in the evaluation of atypical exemplars and their social groups. *Journal of Experimental Social Psychology*, *37*, 386–397. doi:10.1006/jesp.2000.1459

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology*, *55*, 726–737.

Bonilla-Silva, E. (2003). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. New York, NY: Rowman and Littlefield.

Bosson, J. K., Swann, W. B., Jr, & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631–643. doi:10.1037///0022-3514.79.4.631

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*, 330–350. doi:10.1177/1088868312440047

Carlston, D. (2010). Models of implicit and explicit mental representation. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 38–61). New York, NY: Guilford Press.

Chapman, M. V, Hall, W. J., Lee, K., Colby, R., Coyne-Beasley, T., Day, S., & Thomas, T. (in press). Making a difference in medical trainees’ attitudes toward Latino patients: A pilot study of an intervention to modify implicit and explicit attitudes. *Social Science & Medicine*. doi:10.1016/j.socscimed.2017.05.013

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583. doi:10.1016/0169-2070(89)90012-5

Cole, L. W., & Farrell, C. (2006). Structural racism, structural pollution and the need for a new paradigm. *Washington University Journal of Law & Policy*, *20*, 265–282. Retrieved from http://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1247&context=law_journal_law_policy

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428. doi:10.1037/0033-295X.82.6.407

Cooley, E., & Payne, B. K. (2017). Using groups to measure intergroup prejudice. *Personality and Social Psychology Bulletin*, *43*, 46–59. doi:10.1177/0146167216675331

Cooley, E., Payne, B. K., Loersch, C., & Lei, R. (2015). Who owns implicit attitudes? Testing a metacognitive perspective. *Personality and Social Psychology Bulletin*, *41*, 103–115. doi:10.1177/0146167214559712

Cooley, E., Payne, B. K., & Phillips, K. J. (2014). Implicit bias and the illusion of conscious ill will. *Social Psychological and Personality Science*, *5* (4), 500–507. doi:10.1177/1948550613506123

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*, 1332–1346.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170. doi:10.1111/1467-9280.00328

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion*, *9*, 585–591. doi:10.1037/a0015961

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814. doi:10.1037/0022-3514.81.5.800

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342–353. doi:10.1111/spc3.12111

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*, 1267–1278. doi:10.1016/j.jesp.2012.06.003

Devos, T. (2008). Implicit attitudes 101: Theoretical and empirical insights. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (pp. 61–84). New York, NY: Psychology Press.

Dunham, Y., Baron, A., & Banaji, M. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, *77*, 1268–1281. doi:10.1111/j.14678624.2006.00933.x

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*, 248–253. doi:10.1016/j.tics.2008.04.006

Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological Science*, *24*, 860–868. doi:10.1177/0956797612463081

Farkas, I., Helbing, D., & Vicsek, T. (2002). Mexican waves in an excitable medium. *Nature*, *419*(6103), 131. doi:10.1038/419131a

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Fazio, R. H. (2009). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637. doi:10.1521/soco.2007.25.5.603

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327. doi:10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*, 339–357. doi:10.1016/0022-1031(82)90058-0

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology*, *51*, 505–514. doi:10.1037/0022-3514.51.3.505

Feagin, J. R. (2006). *Systemic racism: A theory of oppression*. New York, NY: Taylor & Francis Group.

Feagin, J. R., & Feagin, C. (1986). *Discrimination American style: Institutional racism and sexism* (2nd ed.). Malabar, FL: Krieger.

Ferguson, M. J., & Bargh, J. A. (2003). The constructive nature of automatic evalution. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 169–188). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Freeman, A. D. (1978). Legitimizing racial discrimination through antidiscrimination law: A critical review of supreme court doctrine. *Minnesota Law Review*, *62*, 1049–1119. Retrieved from http://www.dariaroithmayr.com/pdfs/assignments/Freeman,LegitimizingRacialDiscrimination.pdf

Fryberg, S. A., Markus, H. R., Oyserman, D., & Stone, J. M. (2008). Of warrior chiefs and Indian princesses: The psychological consequences of American Indian mascots. *Basic and Applied Social Psychology*, *30*, 208–218. doi:10.1080/01973530802375003

Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 7. doi:10.1038/075450a0

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370–377. doi:10.1016/j.jesp.2006.12.004

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious?. *Consciousness and Cognition*, *15*, 485–499.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*, 300–312. doi:10.1177/0146167216684131

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Gee, G. C., & Ford, C. L. (2011). Structural racism and health inequities: Old issues, new directions. *Du Bois Review*, *8*, 115–132. doi:10.1017/S1742058×11000130

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. doi:10.1037/0022-3514.60.4.509

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27. doi:10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561. doi:10.1037/pspa0000016

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945–968. doi:10.15779/Z38GH7F

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41. doi:10.1037/a0015575

Hahn, A., & Gawronski, B. (2014). Do implicit evaluations reflect unconscious attitudes? *Behavioral and Brain Sciences*, *37*(1), 28–29.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology General*, *143*, 1369–1392. doi:10.1037/a0035028

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology*, *42*, 259–272. doi:10.1016/j.jesp.2005.04.006

Hawkins, C. B., & Nosek, B. A. (2012). Motivated independence?: Implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin*, *38*, 1437–1452. doi:10.1177/0146167212452313

Hehman, E., Flake, J.K., & Calanchini, J. (in press). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science.*

Heiphetz, L., Spelke, E. S., & Banaji, M. R. (2013). Patterns of implicit and explicit attitudes in children and adults: Tests in the domain of religion. *Journal of Experimental Psychology: General*, *142*, 864–879. doi:10.1037/a0029714

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles*. New York, NY: Guilford Press.

Higgins, E. T., King, G. A., & Marvin, G. H. (1982). Individual construct accessibility and subjective impressions and recall. *Journal of Personality and Social Psychology*, *43*, 35–47.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154. doi:10.1016/S0022-1031(77)80007-3

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385. doi:10.1177/0146167205275613

Huntsinger, J. R., & Sinclair, S. (2010). When it feels right, go with it: Affective regulation of affiliative social tuning. *Social Cognition*, *28*, 290–305. doi:10.1521/soco.2010.28.3.290

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*, 1–27. doi:10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*, 586–602.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69. doi:10.1016/j.riob.2009.10.001

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788. doi:10.1037/0022-3514.81.5.774

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32. doi:10.1037/0022-3514.91.1.16

Kassin, S., Fein, S., & Markus, H. R. (2011). *Social psychology* (8th ed.). Belmont, CA: Cengage Learning.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280–290. doi:10.1037/h0074049

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871–888. doi:10.1037//0022-3514.78.5.871

Kihlstrom, J. F. (2004). Implicit methods in social psychology. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 195–212). Thousand Oaks, CA: Sage.

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*, 315–330. doi:10.1111/spc3.12023

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., & Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765–1785. doi:10.1037/a0036260

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., & Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*, 1001–1016. doi:10.1037/xge0000179

LaVeist, T. (1989). Linking residential segregation to the infant mortality disparity in U.S. cities. *Sociology and Social Research*, *73*(2), 90–94.

Lee, K. M., Lindquist, K. A., & Payne, B. K. (in press). Constructing bias: Conceptualization breaks the link between implicit bias and fear of Black Americans. *Emotion*.

Leitner, J.B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine*, *170*, 220–227. doi:10.1016/j.socscimed.2016.10.007

Lepore, L., & Brown, R. (2002). The role of awareness: Divergent automatic stereotype activation and implicit judgment correction. *Social Cognition*, *20*(4), 321–351. doi:10.1521/soco.20.4.321.19907

Loersch, C., & Payne, B. K. (2012). On mental contamination: The role of (mis)attribution in behavior priming. *Social Cognition*, *30*(2), 241–252. doi:10.1521/soco.2012.30.2.241

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin*, *55*, 337–372. doi:10.1037/h0042344

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855. doi:10.1037//0022-3514.81.5.842

Lukachko, A., Hatzenbuehler, M. L., & Keyes, K. M. (2014). Structural racism and myocardial infarction in the United States. *Social Science & Medicine*, *103*, 42–50. doi:10.1016/j.socscimed.2013.07.021

Lundberg, K. B., & Payne, B. K. (2014). Decisions among the undecided: Implicit attitudes predict future voting behavior of undecided voters. *PLoS ONE*, *9*(1), e85680. doi:10.1371/journal.pone.0085680

Mandel, H., & Semyonov, M. (2005). Family policies, wage structures, and gender gaps: Sources of earnings inequality in 20 countries. *American Sociological Review*, *70*, 949–967. Retrieved from http://www.jstor.org/stable/4145401

Marini, M., Sriram, N., Schnabel, K., Maliszewski, N., Devos, T., Ekehammar, B., & Nosek, B. A. (2013). Overweight people have low levels of implicit weight bias, but overweight nations have high levels of implicit weight bias. *PLoS ONE*, *8*(12), e83543. doi:10.1371/journal.pone.0083543

McClelland, J. L. (2000). Connectionist models of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 583–596). New York, NY: Oxford University Press.

McConnell, A. R., Dunn, E. W., Austin, S. N., & Rawn, C. D. (2011). Blind spots in the search for happiness: Implicit attitudes and nonverbal leakage predict affective forecasting errors. *Journal of Experimental Social Psychology*, *47*, 628–634. doi:10.1016/j.jesp.2010.12.018

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning. Memory, and Cognition*, *18*(6), 1155–1172. doi:10.1037/0278-7393.18.6.1155

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist1*, *44*(12), 1469.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*, 1106–1115. doi:10.1177/0956797614524255

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*, 512–523. doi:10.1177/0146167210362789

Miller, C. T., Varni, S. E., Solomon, S. E., DeSarno, M. J., & Bunn, J. Y. (2016). Macro-level implicit HIV prejudice and the health of community residents with HIV. *Health Psychology*, *35*, 807–815. doi:10.1037/hea0000314

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19*(4), 395–417. doi:10.1521/soco.19.4.395.20759

Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, *87*, 817–831. doi:10.1037/0022-3514.87.6.817

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology General*, *134*(4), 565–584. doi:10.1037/0096-3445.134.4.565

Nosek, B. A. (2007). Implicit–explicit relationships. *Psychological Sciences*, *16*, 65–69.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York, NY: Psychology Press.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*(4), 552–594. doi:10.1037/0096-3445.134.4.565

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test. *Experimental Psychology*, *54*, 14–29. doi:10.1027/1618-3169.54.1.14

Nosek, B. A., Smyth, F. L, Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88. doi:10.1080/10463280701489053

Nosek, B. A., Smyth, F. L., Sriram, N., Linder, N. M., Devos, T., Ayala, A., & Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593–10597. doi:10.1073/pnas.0809921106

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667. doi:10.1037/0022-3514.86.5.653

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421–433. doi:10.1177/0146167205284004

Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science & Medicine*, *181*, 191–198. doi:10.1016/j.socscimed.2017.03.036

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192. doi:10.1037/a0032734

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, *108*, 562–571. doi:10.1037/pspa0000023

Page, S. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192. doi:10.1037///0022-3514.81.2.181

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*, 488–503. doi:10.1037/0022-3514.89.4.488

Payne, B. K. (2006). Weapon bias: Split-second decisions and unintended stereotyping. *Current Directions in Psychological Science*, *15*(6), 287–291. doi:10.1111/j.1467-8721.2006.00454.x

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*, 16–31. doi:10.1037/0022-3514.94.1.16

Payne, B. K., & Cameron, C. D. (2013). Implicit social cognition and mental representation. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 220–238). New York, NY: Oxford University Press.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293. doi:10.1037/0022-3514.89.3.277

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology*, *46*, 367–374. doi:10.1016/j.jesp.2009.11.001

Payne, B. K., & Lundberg, K. B. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, *8*, 672–686. doi:10.1111/spc3.12148

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, *25*, 609–642. doi:10.1521/soco.2007.25.5.657

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*, 21–41. doi:10.1037/0022-3514.90.1.21

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*(5), 729–738. doi:10.1162/089892900562552

Powell, J. A. (2008). Structural racism: Building upon the insights of John Calmore. *North Carolina Law Review*, *86*, 791–816. doi:10.1525/sp.2007.54.1.23

Quillian, L. (2008). Does unconscious racism exist? *Social Psychology Quarterly*, *71*, 6–11.

Rae, J. R., Newheiser, A., & Olson, K. R. (2015). Exposure to racial outgroups and implicit race bias in the United States. *Social Psychological and Personality Science*, *6*, 535–543. doi:10.1177/1948550614567357

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44* (2), 386–396. doi:10.1016/j.jesp.2006.12.008

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385. doi:10.1037/0033-295X.95.3.385

Rholes, W. S., & Pryor, J. B. (1982). Cognitive accessibility and cognitive attributions. *Personality and Social Psychology Bulletin*, *8*, 719–729. doi:10.1177/0146167282084019

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, *67*, 439–463. doi:10.1146/annurev-psych-010213-115115

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814. doi:10.1037/0278-7393.21.4.803

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Rudman, L. A. (2004a). Social justice in our minds, homes, and society: The nature, causes, and consequences of implicit bias. *Social Justice Research*, *17*, 129–142. doi:10.1023/B:SORE.0000027406.32604.f6

Rudman, L. A. (2004b). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 80–83. doi:10.1111/j.0963-7214.2004.00279.x

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, *17*(4), 437–465. doi:10.1521/soco.1999.17.4.437

Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes & Intergroup Relations*, *5*, 133–150. doi:10.1177/1368430202005002541

Rudman, L. A., Phelan, J. E., & Heppen, J. (2007). Developmental sources of implicit attitudes. *Personality and Social Psychology Bulletin*, *33*, 1700–1713. doi:10.1177/0146167207307487

Rydell, R.J., & McConnell, A.R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995–1008. doi:10.1037/0022-3514.91.6.995

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition. *European Journal of Psychological Assessment*, *24*(4), 210–217. doi:10.1027/1015-5759.24.4.210

Schwarz, N. (2006). Attitude research: Between Ockam's razor and the fundamental attribution error. *Journal of Consumer Research*, *33*, 19–21. doi:10.1086/504124

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*, 638–656. doi:10.1521/soco.2007.25.5.638

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513–523.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychological Science*, *19*, 717–723. doi:10.1111/j.1467-9280.2008.02147.x

Smith, C. D. (2009). Deconstructing the pipeline: Evaluating school-to-prison pipeline equal protection cases through a structural racism framework. *Fordham Urban Law Journal*, *36*, 1009–1049.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, *70*, 893.

Spalding, L. R., & Hardin, C. D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science*, *10*(6), 535–539. doi:10.1111/1467-9280.00202

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672. doi:10.1037/0022-3514.37.10.1660

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgment. *Journal of Personality and Social Psychology*, *38*, 841–856. doi:10.1037/0022-3514.38.6.841

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*, 1332–1345. doi:10.1177/0146167208321269

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science*, *20*, 164–168. doi:10.1111/j.1467-9280.2009.02274.x

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Doubleday Books.

Tetlock, P. E. & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York, NY: Crown.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*, 1027–1042. doi:10.1037/a0022308

Turner, R. N., & Crisp, R. J. (2010). Imagining intergroup contact reduces implicit prejudice. *British Journal of Social Psychology*, *49*, 129–142. doi:10.1348/014466609×419901

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*, 474–480. Retrieved from http://www.jstor.org/stable/40064251

van Dijk, T. A. (1989). Mediating racism: The role of the media in the reproduction of racism. In R. Wodak (Ed.), *Language, power, and ideology: Studies in political discourse* (pp. 199–226). Philadelphia, PA: John Benjamins Publishing Company.

Webb, T. L., Sheeran, P., & Pepper, J. (2012). Gaining control over responses to implicit attitude tests: Implementation intentions engender fast responses on attitude-incongruent trials. *British Journal of Social Psychology*, *51*, 13–32. doi:10.1348/014466610×532192

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias correction. *Advances in Experimental Social Psychology*, *29*, 141–208. doi:10.1016/S0065-2601(08)60017-9

Wellman, D. (2007). Unconscious racism, social cognition theory, and the legal intent doctrine: The neuron fires next time. In H. Vera & J. R. Feagin (Eds.), *Handbook of the sociology of racial and ethnic relations* (pp. 39–65). New York, NY: Springer. doi:10.1007/978-0-387-70845-4_4

Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports*, *116*, 404–416.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126. doi:10.1037/0033-295X.107.1.101

Zerhouni, O., Rougier, M., & Muller, D. (2016). “Who (really) is Charlie?”: French cities with lower implicit prejudice toward Arabs demonstrated larger participation rates in Charlie Hebdo rallies. *International Review of Social Psychology*, *29*, 69–76, doi:10.5334/irsp.50