

August 2018

Complications in Predicting Intergroup Behavior from Implicit Biases:
One Size Does Not Fit All

Galen V. Bodenhausen and Christopher D. Petsko
Northwestern University

Chapter to appear in:

Krosnick, J. A., Stark, T. H., & Scott, A. L. (Eds.). *The Cambridge handbook of implicit bias and racism*.

Cambridge, England: Cambridge University Press.

Address correspondence to:

Galen V. Bodenhausen
Department of Psychology
Northwestern University
2029 Sheridan Road
Evanston, IL 60208-2710

Email: galen@northwestern.edu
Phone: 847-467-3887

Abstract

Scholars have long recognized that successful prediction of behavior on the basis of explicit attitudes depends on the correspondence between the attitude measure and the focal behavior. Fishbein and Ajzen (2010) argued that behaviors vary in terms of their action, target, context, and time, and that the prediction of specific behaviors is greatly enhanced when explicit attitude measures reflect these features of the to-be-predicted behavior. We argue that the same principle applies in the case of predicting behavior from implicit attitudes, and we review relevant evidence relating to each of Fishbein and Ajzen's parameters. Special attention is paid to the target parameter, given increasing awareness of the intersectional nature of bias. A global race bias may not extend equally to all members of a particular racial identity, and cross-cutting factors such as gender, age, or sexuality may qualify the extent to which global measures of race bias predict discriminatory behavior toward particular individuals.

KEYWORDS: implicit bias, discrimination, attitude-behavior relation, intersectionality, situated social cognition

Complications in Predicting Intergroup Behavior from Implicit Biases: One Size Does Not Fit All

Unlike beliefs, attitudes, and intentions, behaviors are observable events. Any observation of such an event must take place in a certain context and at a given point in time. In addition, most behaviors are directed at some target. It is therefore useful to think of a behavior as composed of four elements: the action performed, the target at which the action is directed, the context in which it is performed, and the time at which it is performed.

Fishbein & Ajzen (2010, p. 29)

Implicit bias has been a topic of great fascination to psychologists and lay people alike. This interest hinges to a substantial degree on the potential connections between automatic (and frequently unwanted) reactions in our minds and subsequent discriminatory behavior. The decades-old literature examining when and how social attitudes are related to behavior has obvious relevance to this possibility and thus provides a rich context for theoretical development. One primary, overarching implication of this extensive literature is that there is little reason to suppose that there will generally be a strong relation between measures of attitudes and behavior. Indeed, the strength of the attitude-behavior relation has regularly been found to be contingent on (a) important moderator variables, such as attitude accessibility (Fazio, Chen, McDonel, & Sherman, 1982), and (b) the degree to which attitudinal and behavioral measures are specified in parallel ways (the *compatibility principle*, Ajzen & Fishbein, 1977). Yet, applying these insights to the case of implicit attitudes has been somewhat inconsistent. To be sure, the moderator-variable approach has been systematically examined, for example in research on implicit bias inspired by the MODE model (Fazio & Olson, 2014). However, concepts and findings underlying the compatibility principle have been less focal in this literature (cf. Karpinsky & Hilton, 2001). Here, we discuss several distinct complications that can emerge when trying to predict discriminatory behavior from implicit measures of bias—complications that are signaled in the dimensions of compatibility summarized in the quote above. Exploring the (mis)alignment of implicit measures with discriminatory behavior in terms of action, target, context, and time affords us an opportunity to examine a variety of important topics, including intersectional biases and social affordances for bias expression.

In the most general terms, the compatibility principle holds that robust prediction of behavior from attitudes requires that both components be assessed at the same level of generality or specificity. A very general or abstract evaluative stance may be poor at predicting any specific, single action related to that attitude, but it might be quite good at predicting the mean (or the sum) of multiple, specific actions having relevance to the attitude in question. This kind

of aggregated index reflects a more general pattern of behavior and thus better corresponds to a general attitude. Conversely, a more specific attitude will be a good predictor of individual actions, provided that the attitude and behavior are aligned in terms of action, target, context, and time. Weigel and colleagues provided comprehensive support for these claims in the domain of explicit attitudes by showing that: (a) a general measure of environmental attitudes was a relatively weak predictor of any single 1 of 14 different pro-environmental behaviors, but it was a robust predictor of the overall behavior pattern, defined by combining all 14 actions into a composite, multiple-act criterion (Weigel & Newman, 1976); and (b) a specifically aligned attitude measure was a much stronger predictor of a specific pro-environmental action than were more general kinds of environmental attitudes (Weigel, Vernon, & Tognacci, 1974). Along similar lines Siegel, Navarro, Tan, and Hyde (2014) found that specifically aligned attitudes explained at least 70% more variance in organ donation decisions, compared to general attitudes.

Can these kinds of findings offer any valuable insights for researchers interested in the relation between implicit attitudes and behavior? The question is timely, given claims that implicit measures like the Implicit Association Test (IAT; Nosek, Greenwald, & Banaji, 2007) offer, at best, only weak prediction of racial/ethnic discrimination (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Outside the domain of intergroup bias, there have been some studies supportive of the importance of compatibility issues in boosting the ability of implicit measures to predict clinical symptoms (Gschwendner, Hofmann, & Schmitt, 2008) as well as consumer choice (Richetin, Perugini, Prestwich, & O’Gorman, 2007). Within the intergroup domain, some suggestive evidence regarding the importance of predictor-criterion compatibility is provided by a recent meta-analysis of research investigating the relation between the IAT and intergroup behavior (Kurdi et al., 2019). In this analysis, the only theoretical variable that was found to moderate the strength of the association between the IAT and intergroup behavior was the degree of compatibility between the measures. In this meta-analytic work, compatibility was coded after the fact, rather than being integrated systematically into a particular research program as, for example, in the previously mentioned research of Weigel and colleagues. The latter kind of research would be especially valuable in appraising the relevance of compatibility issues in the domain of implicit attitudes.

In the remainder of this chapter, we review past research bearing on the potential importance of compatibility issues in the case of implicit bias measures. We organize the discussion around the four dimensions of compatibility enumerated in the theorizing of Fishbein and Ajzen (2010). *Action compatibility* concerns the possibility that different classes or types of behavior differ in their degree of alignment with various kinds of implicit reactions. *Target compatibility* concerns the degree to which the referent of the implicit measure and the referent of the focal action are the same. *Context compatibility* concerns the degree to which the context in which a measured behavior occurs affords expression of the focal implicit attitude. Finally, under the rubric of *time compatibility* we consider how temporal dynamics may modify the relation between implicit attitudes and overt behaviors. In each case, we describe how these

issues can complicate the behavioral expression of bias and thus undermine expectations that there should be a strong, general relationship between implicit bias and intergroup behavior.

Complication #1: Action Compatibility

The idea that implicit bias measures might predict some kinds of behavior better than others was initially highlighted in research by Dovidio, Kawakami, and their colleagues (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). Following the logic of dual-process models of cognition, these scholars argued that implicit bias measures should be well suited to predict relatively spontaneous and hard-to-control forms of behavior (e.g., nonverbal reactions), but they should be less suitable for predicting relatively controlled, deliberated responses (e.g., explicit evaluative judgments). Although this proposition was supported across a number of experiments (for a review, see Perugini, Richetin, & Zogmaister, 2010), theoretical considerations raise some uncertainty regarding the consistency of this hypothesized dissociation pattern. In particular, prominent models of attitude-related cognition propose that under many circumstances, rapid, relatively automatic forms of bias can indeed influence controlled responses by biasing the content of an individual's more effortful thinking (i.e., biased elaboration; see Chaiken & Maheswaran, 1994; Petty, Schumann, Richman, & Strathmann, 1993). Consistent with the possibility that implicit biases might be able to influence both relatively automatic and more controlled kinds of behavior, the meta-analysis conducted by Kurdi et al. (2019) found that the overall strength of the relation between the IAT and intergroup behavior was not moderated by the controllability of the behavioral response. However, the overall strength of that (unmoderated) relation remains modest (while certainly not trivial), and one can ask if there are other moderators that can identify conditions that produce a stronger relation.

A second approach to examining action compatibility focuses on the idea that different forms of implicit bias may be differentially connected to behavior. Amodio and Devine (2006) argued that implicit *attitude* measures capture affective processes and thus should be attuned to consummatory behaviors, whereas implicit *stereotyping* measures capture cognitive/semantic processes and thus should be attuned to instrumental behaviors. If implicit attitudes are automatic affective reactions (e.g., Gawronski & Bodenhausen, 2006; 2011), it stands to reason that they should be especially connected to behaviors that are affect-driven—for example, behaviors reflecting approach or avoidance tendencies, social distancing, etc. (see Elliot, Eder, & Harmon-Jones, 2013). Other forms of discrimination, however, may depend on stereotypic attributes that are semantically associated with a category, rather than on global affective responses (e.g., Bodenhausen, Macrae, & Garst, 1998). For example, to the extent that they hold common social stereotypes, people might be biased against hiring a gay man as a school teacher, relative to other candidates, but be biased in favor of hiring a gay man as a decorator; overall feelings toward gay men per se would be inadequate to capture this pattern of stereotype-based discrimination. Here, too, meta-analytic attempts to compare the relative power of implicit stereotypes versus implicit attitudes in predicting discrimination have failed to produce any evidence of moderation (Oswald

et al., 2013). However, this analysis simply coded whether a stereotype or attitude measure was used, irrespective of the degree of conceptual fit (e.g., semantic overlap) between the predictor and criterion. Research directly varying this factor could prove to be illuminating. For example, implicit racial stereotypes concerning dangerousness may better predict discrimination in a police simulation task, compared to a personnel selection simulation, whereas implicit racial stereotypes concerning laziness may show the converse pattern.

Clearly more work should be devoted to systematic, a priori tests of hypotheses concerning “bias-action fit”—the degree to which an implicit measure overlaps with features of the discriminatory action of interest. Given the attention that has already been devoted to it in the literature, we propose adding a third type of implicit bias to Amodio and Devine’s (2006) distinction between implicit attitudes and implicit stereotypes: implicit identification. In light of this addition, three hypotheses concerning bias-action fit are suggested: (1) Measures of implicit attitudes should produce stronger prediction of affect-infused forms of behavior (such as social distancing) than other kinds of implicit measures do; (2) measures of relevant implicit stereotypes should produce stronger prediction of descriptively specific forms of discrimination than other kinds of implicit measures do; and (3) measures of implicit identification should produce stronger prediction of normative forms of discrimination (such as ingroup favoritism in resource allocation; Kerr, Ao, Hogg, & Zhang, 2018) compared to other kinds of implicit measures. These possibilities warrant further empirical evaluation.

Complication #2: Target Compatibility

A second dimension of compatibility concerns the degree to which the target of a measured implicit bias corresponds to the target of the discriminatory behavior that is being predicted. A major consideration in this regard concerns whether these two targets are specified at the same level of generality. Returning to a previous example from the explicit attitude literature, whether or not someone will join the Sierra Club is better predicted by their (more specific) attitude toward the Sierra Club than by their (more general) attitude toward environmentalism (Weigel et al., 1974). Research on “social bias,” however, commonly focuses on how general attitudes and stereotypes regarding a particular group affect the way specific, individual group members are treated. For there to be any target compatibility in this context, the target of potential discrimination must be perceived as representative of the target group.

In general, whether or not implicit attitudes are activated depends on the degree to which stimulus input matches memory representations of the attitude object (Gawronski & Bodenhausen, 2006). Measures of implicit bias are often tuned to general, default representations of a group, yet many group members will inevitably fail to fit these default assumptions. For example, if researchers wish to use a race IAT to predict nonverbal hostility toward a Black interaction partner, the degree to which the partner activates generic conceptions of a “Black person” is obviously of critical importance. Yet, racial identities are fuzzy, multifaceted constructs (e.g., Deaux, 2018), and specific individuals can vary substantially in how representative they are of a general concept like “Black person.” Indeed, an extensive literature

documents that anti-Black biases are differentially expressed depending on the degree to which individual targets possess racially prototypic features (for reviews, see Maddox, 2004; Maddox & Perry, 2018). In the next sections, we focus on race-based biases as a case study of prototypicality dynamics in constraining the predictive power of implicit bias measures, but parallel considerations could be raised for other kinds of social bias such as sexism, heterosexism, ageism, etc.

Prototypicality as a Master Moderator of Target Compatibility

Americans tend to have clear images in their minds of how Black Americans look and act, and they have no trouble scaling Black targets on the degree to which they seem racially prototypic. Perceived racial prototypicality reflects a holistic judgment and is made on the basis of *many* target features, including the targets' social roles (e.g., Hinzman & Maddox, 2017), their clothing choices (e.g., Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011), and their psychological traits (Cox & Devine, 2015), just to name a few. One aspect that has received marked attention as a cue to target prototypicality is facial Afrocentricity. Afrocentricity is a visual property of targets' faces defined by the extent to which a person looks stereotypically Black (vs. White) in their complexion and physiognomy. Targets with lighter skin tones, thinner noses, and straighter hair are usually considered more Eurocentric, and targets with darker skin tones, wider noses, and more textured hair are considered more Afrocentric (Ma, Correll, & Wittenbrink, 2015). Despite wide variance in phenotypic Afrocentricity of targets, American perceivers can reliably scale faces on this dimension (Blair, Judd, Sadler, & Jenkins, 2002), and these perceptions have proven to be consequential for targets over and over again.

In the criminal sentencing domain, individuals who look more prototypically Black, or more Afrocentric, receive longer prison sentences than those who look less prototypically Black (Blair, Judd, & Chapleau, 2004). This is true even when researchers control for how many crimes these individuals have committed in the past, and for how severe their current crimes are. In a disturbing examination of court records from the Philadelphia criminal justice system, Eberhardt and colleagues (2006) found that Black men with more prototypically Black faces (as judged by their mug shots) were substantially more likely than men with less prototypic faces to have received the death penalty—even though all the target men in this study had been accused of the same crime. These examples are correlational, but the principle that Black men who seem more prototypic are targeted more often by racial biases holds in experimental contexts as well. For instance, when participants play a video game in which they are asked to shoot at targets who are armed with guns and to not shoot at targets who are unarmed, they show a replicable “shooter bias” involving a greater readiness to mistakenly shoot unarmed Black men than unarmed White men (Correll, Park, Judd, & Wittenbrink, 2002). More recent studies demonstrate that this tendency is exacerbated when the Black targets in question are manipulated to be prototypical vs. non-prototypical (Kahn & Davies, 2010; Ma & Correll, 2011). Target prototypicality moderates targets' racialized outcomes in many other domains besides the domains of violence and criminal prosecution, including educational attainment (Maddox &

Perry, 2018), personnel selection (Harrison & Thomas, 2009), and in earnings (Devaraj, Quigley, & Patel, 2018). The fundamental point is that target prototypicality is a feature that strongly dictates a person's likelihood of encountering bias. As such, it is likely to be a strong determinant of whether or not a general measure of implicit race bias is useful in predicting reactions to a particular individual.

An interesting new wrinkle on the topic of how target prototypicality relates to bias expression concerns the functional form of this relationship. The theoretical considerations considered so far lead to the clear expectation of a linearly increasing function: greater target prototypicality should be associated with greater bias expression. However, Ma, Correll, and Wittenbrink (2018) recently noted that most extant tests of the relation have involved just two levels of prototypicality (i.e., lower versus higher), so a linear effect was the only testable relation. In new studies comparing low, average, and high levels of the racial prototypicality of Black and White men's faces, Ma et al. found evidence for both the expected positive linear function but also for a significant quadratic relationship, with average levels of target prototypicality being associated with the strongest expressions of bias. It thus seems likely that extremely prototypic group members trigger some additional psychological reaction(s), besides more strongly activating the category's cognitive representation, and these reactions in turn can constrain bias expression. Also of note, both the linear and quadratic effects observed in this research were linked to differential expression of negative stereotypes, but not of positive ones. The seemingly distinct effects of racial prototypicality on negative versus positive kinds of bias, as well as the specific factors producing reduced bias expression at high levels of prototypicality, are topics ripe for further investigation.

Prototypicality and the Intersectional Nature of Bias

A social category like "Black people" encompasses a diverse array of individuals, including people of all ages, genders, sexual orientations, and socioeconomic statuses. Yet, when mentally representing this category, do we typically consider this great diversity? In short, no. Instead, our category prototypes are constructed around certain default features on various cross-cutting demographic dimensions. The most prototypic member of many social categories is male (Hamilton, 1991; Bailey, LaFrance, & Dovidio, 2018), heterosexual (Lick & Johnson, 2016), and a young adult (for a review, see Bodenhausen & Peery, 2009). African Americans are also presumed to be of lower socioeconomic status (SES) by default (Cox & Devine, 2015). A consequence of this tendency to build various demographic assumptions into the prototypic image of a Black person is that, when attitudes and stereotypes about "Black people" are measured, they may actually refer primarily to a subset of this category, namely to male, heterosexual, low-SES, young-adult Black people. That is, the de facto bias that is being measured may be more specific than intended. The principle of target compatibility suggests that these measures may be less suitable for predicting discrimination toward Black people who do not fit this default demographic prototype (e.g., Black women, Black senior citizens, etc.). This observation aligns with the current emphasis in prejudice research on the importance of

considering the intersectional nature of social bias (e.g., Kang & Bodenhausen, 2015; Nicolas, de la Fuente, & Fiske, 2017). We next consider what this literature currently reveals about the generalizability of racial bias across these intersections.

Race and gender. Much research has been conducted on the question of how racial bias is attenuated or exacerbated depending on targets' gender. The prototypicality argument sketched out above leads to the expectation that racial biases will generally be more strongly directed toward male targets than toward female targets. This same prediction emerges from the theory of gendered prejudice (McDonald, Sidanius, & Navarrete, 2011), which holds that males' greater size and physical strength (on average) makes them a more salient source of threat and thus a more likely target of race-based anxieties, antipathies, and aspirations for dominance and subjugation. Consistent with this perspective is the fact that racial gaps in incarceration rates are greater among male targets than among female targets and that reports of racial discrimination are more pronounced among Black male respondents than among Black female respondents (Sidanius & Pratto 1999; Veenstra, 2013). Research on automatic forms of race bias has produced converging findings. For example, young children express automatic race biases more toward male than female peers (Perszyk, Lei, Bodenhausen, Richeson, & Waxman, 2019), and the previously discussed racial shooter bias has been found repeatedly for male targets but is absent for female targets (Plant, Goplen, & Kunstman, 2011). Conversely, a perspective-taking intervention designed to change overall race biases produced stronger bias reduction when it focused on male versus female Black targets (Todd & Simpson, 2016).

An alternative perspective argues that having two subordinated identities exacerbates the likelihood of experiencing discrimination, an idea known as the double-jeopardy perspective (e.g., Crenshaw, 1995). In the context of Black women, this implies that racial and gender biases may compound in ways that result a greater likelihood of experiencing discrimination. Which of these perspectives is correct? Are Black men or Black women more likely to be targeted for discrimination? According to intersectional invisibility theory (Purdie-Vaughns & Eibach, 2008), both approaches have validity, but within different spheres of discrimination. If we assume that Black women are indeed less racially prototypic than Black men, this means that they may—in line with the theory of gendered prejudice—be less often targeted by active racial biases that reflect motivations to defend oneself against or dominate the racial outgroup. However, the lower prototypicality of Black women—both in terms of their race and their gender—can make them more vulnerable to passive biases that result from their relative invisibility. These biases include being overlooked, forgotten, or having one's contributions misattributed to others (e.g., Goff, Thomas, & Jackson, 2008; Sesko & Biernat, 2010), phenomena that certainly are disadvantageous in many contexts. Little is known about whether these more passive kinds of biases are related to individual difference measures of implicit attitudes, but it would not be surprising if such a relationship were absent. In contrast, active forms of discrimination that build from negative or competitive thoughts and feelings about an outgroup may very well relate to the strength of implicit biases—biases that are, even if only implicitly, more targeted at men than women.

Race and age. Psychologists have recently explored whether race bias is directed similarly toward targets of various age groups. As noted above, many social group prototypes are presumed young-adult by default. Thus, we might expect racial minorities who are young adults to experience more discrimination than those who are young children or seniors. Some research supports this conclusion. For instance, there is a well-established bias in perceivers' minds to perceive Black faces as emotively angrier than White faces (Hugenberg & Bodenhausen 2003; 2004). However, these studies generally use faces of Black and White young adults. When the faces are systematically manipulated to be of older vs. young adult men, the bias is moderated. The tendency to "see" anger on the faces of Black men occurs more when the targets are more prototypic young-adult men than when they are less prototypic older-adult men (Kang & Chasteen, 2009; Kang, Chasteen, Cadieux, Cary, & Syeda, 2014).

Other findings, however, complicate conclusions about age-based moderation of race biases. For example, it is well established that priming White perceivers with the faces of Black (relative to White) targets increases the speed with which they can recognize criminal objects (Eberhardt, Goff, Purdie, & Davies, 2004; Payne, 2001). These studies, too, have primarily used young adult faces. What happens when researchers use child and older adult targets? Surprisingly, the bias remains robust. Just as young-adult Black men's faces facilitate perceivers' sensitivity to crime objects, so too do young (Todd, Thiem, & Neel, 2016) and older (Lundberg, Neel, Lassetter, & Todd, 2018) male faces. The finding that Black boys can evoke race biases that are quite similar in magnitude to those evoked by young adults could reflect the fact that perceivers tend to overestimate the age of Black boys and thus do not see them as deviating as much from the young-adult group prototype as their same-age White counterparts (Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014). White perceivers are generally less able to determine the age of Black (vs. White) targets shown in photos (Dehon & Brédart, 2001), so the variability in findings may depend at least in part on differences in the clarity and extremity with which age is manipulated in visual stimuli. Alternatively, age could simply be a relatively weak moderator of overall racial prototypicality. In contrast, it appears to play a more pronounced role in moderating gender stereotyping (e.g., Koenig, 2018; Martin, North, & Phillips, 2018).

Race and sexual orientation. Several recent studies have shown that racial bias is differentially exhibited toward heterosexual vs. homosexual Black targets. For example, biases against Black men in personnel selection tend to dissipate when the target men in question are homosexual rather than heterosexual (Pedulla, 2014; Wilson, Remedios, & Rule, 2017). Moreover, whereas people typically stereotype Black men more negatively than White men, this tendency reverses when targets are homosexual (Remedios, Chasteen, Rule, & Plaks, 2011). Some of our own recent data suggest that Black men who are described as homosexual are not only seen as less stereotypically Black, but are also stereotyped as having substantially Whiter characteristics than their heterosexual counterparts (Petsko & Bodenhausen, 2019b). These findings indicate that target sexual orientation, if detectable or known, may play an important role in moderating targets' racial prototypicality and thus the relation between implicit bias and behavior against those targets.

Race and socioeconomic status. A final demographic feature of targets that might moderate how racially prototypic they seem is their ostensible socioeconomic status. This possibility is supported by an impressive analysis of longitudinal data by Penner and Saperstein (2008). In a nationally representative data set involving more than 12,000 Americans, Penner and Saperstein found that one fifth of their sample changed races—from the perspective of door-to-door interviewers—at least once over a 17-year time frame. Moreover, they found that these racial classification changes were not determined randomly, but instead were correlated with whether the respondent told their interviewer that they had moved up or down the socioeconomic ladder. When respondents had moved down the ladder—when they had been recently incarcerated, unemployed, or began living in poverty—interviewers became more likely to categorize them as non-White, even if they had categorized respondents as White at a previous point in time. Freeman et al. (2011) followed up on these findings by showing that a reverse pattern can occur as well: when targets' clothing is experimentally manipulated to be high-SES (i.e., a business suit) rather than low-SES (i.e., a janitorial outfit), perceivers become more likely to categorize them as White and less likely to categorize them as Black. Research examining the implications of these patterns for implicit biases corroborates the importance of the SES-notations of a target's clothing; for example, Kahn and Davies (2017) found that racial bias in the shooter task is exacerbated when targets are depicted wearing stereotypically lower-class clothing (baggy sweatshirt and baseball cap), compared to more middle-class attire (a button-down shirt and tie). The upshot of these findings is that SES serves as an important cue to prototypic Blackness. Thus, higher-SES Black targets may be less often targeted by racial bias than lower-SES Black targets.

Complication #3: Context Compatibility

The research on intersectional identities highlights another way in which compatibility can vary. Because all people possess multifaceted identities, it is an open question whether they will be evaluated in terms of one particular category (such as race) versus a different one (such as gender). Social contexts are likely to moderate the expression of implicit biases to the extent that they differentially influence the salience or relevance of particular social categories (Bodenhausen & Macrae, 1998; Oakes, Turner, & Haslam, 1991; Petsko & Bodenhausen, 2019c). A person high in implicit race bias may well act in racist ways when race is contextually highlighted, but he or she may fail to do so when race is contextually less relevant. For example, when a particular situation makes race salient, Asian women are seen as more skilled and more hireable, compared to when the situation makes gender salient (Rattan, Steele, & Ambady, 2019). Thus, different contexts carry different situational affordances for the expression of race (or other) biases.

In addition to contextual variation in social category relevance, contexts also vary in the incentives they hold for different kinds of cognitive activity. For example, situations can influence accuracy motivation, social-affiliative motivation, or ego-defensive motivation, and these different kinds of motives might be differentially satisfied by expressing versus

suppressing social biases. Situational contexts also enable or constrain opportunities for thinking beyond one's most immediate, automatic reactions. Because the importance of these factors has been extensively analyzed elsewhere (e.g., Chaiken, Liberman, & Eagly, 1989; Fazio, 1990; for multiple reviews, see Sherman, Gawronski, & Trope, 2014), we focus our attention here on issues related to the multiplicity of social identity and the contextual factors that highlight particular social categories within this panoply.

What factors govern the relevance (and usage) of particular social categories? According to Bruner (1957), perceivers use a given category when it is highly accessible to the perceiver (either chronically or situationally) and when it "fits" reality. Oakes and colleagues (1987; 1991) expanded on this idea by arguing that in the social realm, fit comes in one of two forms: comparative fit and normative fit. *Comparative fit* describes the extent to which a social category maximizes the ratio of perceived differences across clusters of people relative to perceived similarities within clusters of people. As an example, suppose a group of people of different races and genders is having an informal debate. If the discussion revealed that the women (whatever their race) agreed with one another in taking a particular position and the men also tended to agree with one another in taking a different position, then gender would provide good comparative fit, whereas race would provide poor comparative fit. As such, gender should be a more salient basis for perceiving and responding to the individuals in question. But if the expressed opinions happened to cluster by race irrespective of gender, then race should become focal and hence be a more potent source of potential bias. *Normative fit* describes the extent to which a social category normatively (i.e., stereotypically) aligns with patterns of observed behavior. For example, the gender of an Asian woman may be more salient if she is observed advocating for better childcare policies, but her racial identity may be more salient if she is advocating for technology innovation (see Craig & Bodenhausen, 2018, for other examples). The key insight of this approach, which is generally referred to as the social identity approach (Abrams & Hogg, 2010), is that situations vary in the extent to which they make certain identities accessible and fitting. The greater the social category fit, the more we can expect biases associated with that category to guide perceivers' behaviors toward targets.

The importance of comparative fit is also highlighted in research examining the salience of race, depending on whether other kinds of groupings provide a better indication of shared goals and likemindedness. Arguing from an evolutionary perspective, Cosmides, Tooby, and Kurzban (2003) proposed that, rather than being "hardwired" to attend to any particular group categorization (such as race), we are sensitized to detecting patterns of cooperation and coaction that reflect coalitional groupings. From this perspective, it is only to the extent that race is perceived to covary with coalitional membership that it should become a salient identity boundary. In a demonstration of this point, Kurzban, Tooby, and Cosmides (2001) showed that categorizing targets by race can be attenuated when perceivers perceive an alternative, nonracial signal of the functional coalitions within an intergroup context. This perspective implies that those perceivers who do chronically activate racial categories are using race as a proxy to infer

intergroup allegiances, oblivious of any other, potentially better fitting bases for defining coalitions.

What about accessibility? As noted above, accessibility of social categories—the ease with which people use them—can vary situationally. It probably comes as no surprise that when people are asked to incidentally consider their own or others' demographic categories, these category memberships become more accessible and are hence more likely to be used (e.g., Rattan et al., 2019). But, another feature of social contexts that influences the accessibility of social categories is the extent to which they are relatively rare and thus attention-grabbing (McGuire & McGuire, 1981; Nelson & Miller, 1995). For example, perceivers who walk into a room with five White people and a single Black person may be especially likely to use race as a social category. Indeed, there are many such reports of *solo status* influencing the accessibility of social categories both for perceivers (e.g., Biernat & Vescio, 1993) and for the person who is occupying the sole position as a perceived ambassador for his or her social group (Sekaquaptewa & Thompson, 2003). Thus, shortages of social category representation in a given context can somewhat ironically make attending to it more probable, potentially exacerbating latent biases against members of that category.

In addition to varying in how much they highlight certain identity dimensions over others, contexts can vary in their relevance to particular kinds of bias. As previously mentioned, stereotypes are only relevant in certain situations, where the stereotypic characteristic has some bearing on how a person is perceived, evaluated, or treated. A stereotype such as “Blacks are athletically talented” should have much more impact in situations where physical performance is relevant than in other contexts, even if the target person being evaluated is equally strongly categorized as Black in all these contexts. As another example, although Blacks are commonly stereotyped in ways that are relevant to their supposed propensity to commit various kinds of violent versus white-collar crimes, gay men are not; thus, a manipulation of race (Black versus White) was found to influence perceptions of accused criminals whereas a cross-cutting manipulation of sexual orientation (gay versus heterosexual) was not (Petsko & Bodenhausen, 2019a). The bias-relevance of context was also investigated by Barden, Maddux, Petty, and Brewer (2004), who found that racial biases were more strongly expressed toward Black targets depicted in a prison context than in a church context. The stereotypic idea that Blacks are dangerous or threatening simply has far less relevance in a church context. Along similar lines, Correll, Wittenbrink, Park, Judd, and Goyle (2011) found that the racial shooter bias was stronger when targets were depicted in dilapidated urban environments compared to more neutral situations.

Collectively, these ideas indicate that implicit biases should be much more likely to predict discriminatory treatment when situations make bias-relevant social identities salient (by virtue of their fit or contextual accessibility), or when contextual features reinforce versus undermine the relevance of the bias in question.

Complication #4: Time Compatibility

Many temporal matters are important when it comes to debates about implicit biases, including issues such as the test-retest reliability of implicit measures (e.g., Rae & Olson, 2018) as well as the temporal durability of induced reductions in implicit biases (e.g., Devine, Forscher, Austin, & Cox, 2012). Here we focus on one particular temporal parameter: the elapsed time between encountering and responding to an outgroup target. Implicit measures are designed to capture rapid, immediate reactions to a particular attitude object. Thus, time compatibility would be maximized when behavioral forms of discrimination also involve relatively rapid, immediate reactions (whether these involve nonverbal behavior, snap judgments, or other kinds of actions). As more time elapses between encountering targets and responding to them, a variety of additional psychological processes can potentially dilute the impact of immediate response biases. As we noted earlier, recent meta-analyses have failed to produce much evidence for a role of behavior controllability in moderating the expression of implicit biases (e.g., Kurdi et al., 2019). Although controllability and processing time are related to one another, they are not isomorphic concepts. Direct examinations of the time course of bias provide a basis for further investigation of this potential moderator of compatibility.

In one particularly relevant set of studies, Kunda, Davies, Adams, and Spencer (2002) examined the time course of participants' activation of racial stereotypes while they watched a video of an African American being interviewed. Stereotype activation was evident very early on (15 seconds after the interview commenced), but after 12 minutes had elapsed, no evidence for stereotype activation was seen. However, when the target being interviewed expressed an opinion that participants disagreed with, stereotype activation was interestingly found to reemerge late in the interview. This pattern indicates that activation of implicit biases is quite dynamic and sensitive to changing features of the situation over time. This work underscores the proposition that implicit measures should best predict rapid, more impulsive forms of discrimination, but they may also have valuable predictive power after subsequent events in an ongoing interaction, if these events engender motivation to reconsider or reinterpret the target's qualities (such as if they express disagreement with the perceiver's views).

Rivers, Sherman, Rees, Reichardt, and Klauer (2019) explored related issues and also found processing time to be a noteworthy moderator of intergroup bias. Specifically, they found that stereotype application was significantly reduced as the time available for processing stimulus information increased. This reduction was attributable to a bias-control process that required time to unfold. Interestingly, and in contrast to the implications of the work of Kunda et al., these researchers found evidence that, while stereotype application decreased as processing time increased, stereotype activation actually continued to increase with additional processing time. In this case it does not seem that the category per se becomes less salient over time but rather that people use available time to implement control strategies to avoid applying the activated stereotypes they have in mind. In any case, this research, like that of Kunda et al., corroborates the idea that the behavioral expression of implicit biases is time-dependent.

Some Implications for Future Research

The principle of compatibility, most fundamentally, holds that predictor variables should be constructed so as to maximize their alignment with the behavior that one wants to predict. Most of the existing implicit measures of race bias seek to assess a global kind of bias that applies broadly toward many or most members of the racial group in question, in a variety of different contexts or circumstances. However, most tests of the predictive power of implicit measures involve only measuring one particular kind of behavior, directed at particular targets in one particular situational context. The likelihood of robust prediction under these circumstances is limited. As such, the predictive value of very general indicators of racial bias would better be evaluated in terms of their relation to more general patterns of intergroup behavior, as assessed in multi-act criteria that collapse across different actions, targets, contexts, and time parameters. More tests of this sort would provide a very useful addendum to debates about whether the degree of predictive precision offered by implicit measures is strong enough to care about (Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015).

Sometimes, of course, researchers are not interested in predicting an individual's general pattern of behavior across diverse contexts but instead want to predict some relatively specific form of discrimination (e.g., over-policing of African Americans in a particular urban setting). The considerations we have reviewed suggest that general measures of implicit bias may not be the ideal choice as a predictor in this situation. Instead, implicit measures that optimize the fit to features of the focal behavior might be constructed. Establishing this fit would involve considering the kind of action involved (e.g., how affect-laden it is, how attuned it is to specific semantic beliefs, how much it reflects normative patterns of action), the kinds of targets who will be encountered (e.g., only young-adult men?), the social contexts to be studied (e.g., how much they highlight or de-emphasize the focal identity), and the time parameters of the behavior (e.g., how much does it involve rapid, first-reactions). Having ready-to-go, all-purpose measures of racial bias is an understandable aspiration, but it may be unrealistic if predicting specific kinds of discrimination is a researcher's goal. Although adapting or creating measures of more specifically compatible implicit biases may require extra effort, the meta-analytic results of Kurdi et al. (2019) provide initial reason to view such efforts as very much worthwhile; in their analysis, the correlation between implicit bias and intergroup behavior rose to $r = .37$ in the relatively few studies ($k = 13$) where the degree of bias-criterion correspondence was high, a substantially stronger relation than the overall correlation of .11.

Particular implicit measures vary in the extent to which they can be readily tailored in ways that increase their compatibility with the behavioral criterion. Target and context factors can be readily modified in the Affect Misattribution Procedure (AMP; Payne & Lundberg, 2014). For example, a race AMP could use as its stimuli photos of racial group members who are either male or female, young or old, in a gritty urban setting or a comfortable suburban one, etc. Because the IAT estimates the strength of associations between pairs of dichotomous response categories, tailoring it may be more logistically complicated. For example, if one wants to

measure an implicit attitude toward Black women in an urban context, one could ask participants to sort images of Black urban women into one category, but then how should the other category be defined? Using this “residual” category as a non-focal category in the Brief IAT (Sriram & Greenwald, 2009) might be one option.

One particular focus of our discussion has been the fact that race biases are not distributed equally across all members of a racial category. An implication of this fact for the measurement of bias is that the particular stimuli that are used to assess race bias will influence the apparent magnitude of the respondents’ biases (e.g., Stepanova, Strube, Clote, & Limes, 2016; Wolsiefer, Westfall, & Judd, 2017). One strategy for dealing with this issue might be to select only highly prototypic group members as stimuli for implicit measures, and another might be to sample broadly from the racial category in question to capture the range of variability and model this variability within data that are collected. The first strategy could be thought of as capturing the associations of the core representation of the racial category, while the second strategy better reflects the category’s full breadth. Another strategy, advocated by Cooley and Payne (2017), is to use images of groups instead of individuals when assessing implicit biases. For tasks like the AMP, which in its original version assessed reactions to photos of specific individuals, having several members of the same racial group appear in any given stimulus image should greatly increase the likelihood that race per se is shaping respondents’ automatic evaluative reactions, rather than idiosyncratic affordances of a given individual’s facial physiognomy or other cross-cutting identities possessed by individual targets. Cooley and Payne report evidence that using group-based stimuli in the race AMP increased its test-retest reliability and predictive validity relative to a traditional AMP that used pictures of various individuals as stimuli.

In the history of psychology, certain ideas crop up repeatedly in different contexts. One such idea is the principle of aggregation (e.g., Epstein, 1983), reflected here in the expectation that a general implicit attitude measure should predict an aggregated pattern of behavior much more effectively than any single action. Another recurring and related idea is the recognition that virtually any particular, specific behavior is going to be multiply determined (e.g., O’Grady, 1982). An implicit bias is only one of many possible influences when members of an outgroup are encountered in a specific context. If we want to maximize the chance to observe a role of implicit bias in predicting behavior under these circumstances, we should make our implicit assessments as compatible as possible with the specific form of discrimination we are interested in. When tailored in this way, explicit attitude measures have proven to be quite meaningfully predictive of specific behaviors (Ajzen & Fishbein, 1973), and the same may be true of implicit attitude measures.

References

- Abrams, D., & Hogg, M. A. (2010). Social identity and self-categorization. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping and discrimination* (pp. 179-193). London: Sage Publications.
- Ajzen, I., & Fishbein, M. (1973). Attitudinal and normative variables as predictors of specific behavior. *Journal of Personality and Social Psychology, 27*, 41-57.
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888-918.
- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652-661.
- Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2018). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review*. <https://doi.org/10.1177/1088868318782848>
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5-22.
- Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology, 29*, 166-202.
- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*, 674-679.
- Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*, 5-25.
- Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer Jr. (Ed.), *Stereotype activation and inhibition: Advances in social cognition* (Vol. 11, pp. 243-256). Mahwah, NJ: Erlbaum.
- Bodenhausen, G. V., Macrae, C. N., & Garst, J. (1998). Stereotypes in thought and deed: Social-cognitive origins of intergroup discrimination. In C. Sedikides, J. Schopler, & C. A. Insko (Eds.), *Intergroup cognition and intergroup behavior* (pp. 311-335). Mahwah, NJ: Erlbaum.
- Bodenhausen, G. V., & Peery, D. (2009). Social categorization and stereotyping in vivo: The VUCA challenge. *Social and Personality Psychology Compass, 3*, 133-151.
- Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*, 123-152.
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-473.
- Cooley, E., & Payne, B. K. (2017). Using groups to measure intergroup prejudice. *Personality and Social Psychology Bulletin, 43*, 46-59.

- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*, 1314-1329.
- Correll, J., Wittenbrink, B., Park, B., Judd, C. M., & Goyle, A. (2011). Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of Experimental Social Psychology, 47*, 184-189.
- Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7*, 173-179.
- Cox, W. T. L., & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PLoS ONE, 10*, 1-27.
- Craig, M. A., & Bodenhausen, G. V. (2018). Category (non)fit modulates extrapolative stereotyping of multiply categorizable targets. *Social Cognition, 36*, 559-588.
- Crenshaw, K. W. (1995). Mapping the margins: Intersectionality, identity politics and violence against women of color. In K. Crenshaw, N. Gotanda, G. Peller, & K. Thomas (Eds.), *Critical race theory: The key writings that formed the movement* (pp. 357-383). New York: New Press.
- Deaux, K. (2018). Ethnic/racial identity: Fuzzy categories and shifting positions. *The ANNALS of the American Academy of Political and Social Science, 677*, 39-47.
- Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. *Perception, 30*, 1107-1113.
- Devaraj, S., Quigley, N. R., & Patel, P. C. (2018). The effects of skin tone, height, and gender on earnings. *PloS ONE, 13*, 1.
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*, 1267-1278.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510-540.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383-386.
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876-893.
- Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach-avoidance motivation and emotion: Convergence and divergence. *Emotion Review, 5*, 308-311.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51*, 360-392.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology, 23*, 75-109.

- Fazio, R. H., Chen, J. M., 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.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude—behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 155-171). New York: Guilford.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York: Psychology Press.
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS One, 6*, 1-10.
- 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.
- Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59-127.
- Goff, P. A., Jackson, M. C., Di Leone, D. B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology, 106*, 526-545.
- Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). “Ain’t I a woman?”: Towards an intersectional approach to person perception and group-based harms. *Sex Roles, 59*, 392-403.
- 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.
- Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Convergent and predictive validity of implicit and explicit anxiety measures as a function of specificity similarity and content similarity. *European Journal of Psychological Assessment, 24*, 254-262.
- Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People = male, male = people. *Psychology of Women Quarterly, 15*, 393-402.
- Harrison, M. S., & Thomas, K. M. (2009). The hidden prejudice in selection: A research investigation on skin color bias. *Journal of Applied Social Psychology, 39*, 134-168.
- Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology, 70*, 95-105.
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640-643.
- Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science, 15*, 342-345.
- Kahn, K. B., & Davies, P. G. (2010). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations, 14*, 569-580.

- Kahn, K. B., & Davies, P. G. (2017). What influences shooter bias? The effects of suspect race, neighborhood, and clothing on decisions to shoot. *Journal of Social Issues, 73*, 723-743.
- Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology, 66*, 547-574.
- Kang, S. K., & Chasteen, A. L. (2009). Beyond the double-jeopardy hypothesis: Assessing emotion on the faces of multiply-categorizable targets of prejudice. *Journal of Experimental Psychology, 45*, 1281-1285.
- Kang, S. K., Chasteen, A. L., Cadieux, J., Cary, L. A., & Syeda, M. (2014). Comparing young and older adults' perceptions of conflicting stereotypes and multiply-categorizable targets. *Psychology and Aging, 29*, 469-481.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*, 774-788.
- Kerr, N. L., Ao, X., Hogg, M. A., & Zhang, J. (2018). Addressing replicability concerns via adversarial collaboration: Discovering hidden moderators of the minimal intergroup discrimination effect. *Journal of Experimental Social Psychology, 78*, 66-76.
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology, 9*, 1086.
- Kunda, Z., Davies, P. G., Adams, B. D., & Spencer, S. J. (2002). The dynamic time course of stereotype activation: Activation, dissipation, and resurrection. *Journal of Personality and Social Psychology, 82*, 283-299.
- Kurdi, B., Seitchik, A., Axt, J., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist, 74*, 569-586.
- Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*, 15387-15392.
- Lick, D. J., & Johnson, K. L. (2016). Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments. *Journal of Personality and Social Psychology, 110*, 801-817.
- Lundberg, G. J. W., Neel, R., Lassetter, B., & Todd, A. R. (2018). Racial bias in implicit danger associations generalizes to older male targets. *PLoS One, 13*, 1-17.
- Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47*, 391-396.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*, 1122-1135.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2018). The effects of category and physical features on stereotyping and evaluation. *Journal of Experimental Social Psychology, 79*, 42-50.
- Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review, 8*, 383-401.
- Maddox, K. B., & Perry, J. M. (2018). Racial appearance bias: Improving evidence-based policies to address racial disparities. *Policy Insights from the Behavioral and Brain Sciences, 5*, 57-65.

- Martin, A. E., North, M. S., & Phillips, K. W. (2019). Intersectional escape: Older women elude agentic prescriptions more than older men. *Personality and Social Psychology Bulletin, 45*, 342-359.
- McDonald, M. M., Sidanius, J., & Navarrete, C. D. (2011). Developing a theory of gendered prejudice: An evolutionary and social dominance perspective. In R. M. Kramer, G. J. Leonardelli, & R. W. Livingston (Eds.), *Social cognition, social identity, and intergroup relations* (pp. 189–220). New York: Psychology Press.
- McGuire, W. J., & McGuire, C. V. (1981). The spontaneous self-concept as affected by personal distinctiveness. In M. Lynch, A. A. Norem-Hebeison, & K. Gergen (Eds.), *Self-concept: Advances in theory and research* (pp. 147-171). New York: Ballinger.
- Nelson, L. J., & Miller, D. T. (1995). The distinctiveness effect in social categorization: You are what makes you unusual. *Psychological Science, 6*, 246-249.
- Nicolas, G., de la Fuente, M. D., & Fiske, S. T. (2017). Mind the overlap in multiple categorization: A review of crossed categorization, intersectionality, and multiracial perception. *Group Processes & Intergroup Relations, 20*, 621-631.
- 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: Psychology Press.
- Oakes, P. J. (1987). The salience of social categories. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell, *Rediscovering the social group* (pp. 117-141). Oxford, UK: Blackwell.
- Oakes, P. J., Turner, J. C., Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology, 30*, 125-144.
- O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin, 92*, 766-777.
- 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.
- 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.
- 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.
- Payne, K., & Lundberg, K. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8*, 672-686.
- Pedulla, D. S. (2014). The positive consequences of negative stereotypes: Race, sexual orientation, and the job application process. *Social Psychology Quarterly, 77*, 75-94.
- Penner, A. M., & Saperstein, A. (2008). How social status shapes race. *Proceedings of the National Academy of Sciences, 105*, 19628-19630.

- Perszyk, D. R., Lei, R. F., Bodenhausen, G. V., Richeson, J. A., & Waxman, S. R. (2019). Bias at the intersection of race and gender: Evidence from preschool-aged children. *Developmental Science*, *22*, e12788.
- Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 255-278). New York: Guilford.
- Petsko, C. D., & Bodenhausen, G. V. (2019a). Race-crime congruency effects revisited: Do we take defendants' sexual orientation into account? *Social Psychological and Personality Science*, *10*, 73-81
- Petsko, C. D., & Bodenhausen, G. V. (2019b). Racial stereotyping of gay men: Can a minority sexual orientation erase race? *Journal of Experimental Social Psychology*, *83*, 37-54.
- Petsko, C. D., & Bodenhausen, G. V. (2019c). Multifarious person perception: How social perceivers manage the complexity of intersectional targets. *Social and Personality Psychology Compass*, e12518. doi: 10.1111/spc3.12518
- Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high-and low-elaboration conditions. *Journal of Personality and Social Psychology*, *64*, 5-20.
- Plant, E. A., Goplen, J., & Kunstman, J. W. (2011). Selective responses to threat: The roles of race and gender in decisions to shoot. *Personality and Social Psychology Bulletin*, *37*, 1274-1281.
- Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*, 377-391.
- Rae, J. R., & Olson, K. R. (2018). Test–retest reliability and predictive validity of the Implicit Association Test in children. *Developmental Psychology*, *54*, 308-330.
- Rattan, A., Steele, J., & Ambady, N. (2019). Identical applicant but different outcomes: The impact of gender versus race salience in hiring. *Group Processes & Intergroup Relations*, *22*, 80-97.
- Remedios, J. D., Chasteen, A. L., Rule, N. O., & Plaks, J. E. (2011). Impressions at the intersection of ambiguous and obvious social categories: Does gay + Black = likable? *Journal of Experimental Social Psychology*, *47*, 1312-1315.
- Richetin, J., Perugini, M., Prestwich, A., & O'Gorman, R. (2007). The IAT as a predictor of food choice: The case of fruits versus snacks. *International Journal of Psychology*, *42*, 166-173.
- Rivers, A. M., Sherman, J. W., Rees, H. R., Reichardt, R., & Klauer, K. C. (2019). On the roles of stereotype activation and application in diminishing implicit bias. *Personality and Social Psychology Review*. <https://doi.org/10.1177/0146167219853842>
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*, 68-74.
- Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*, 356-360.
- Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.) (2014). *Dual-process theories of the social mind*. Guilford Publications.

- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Siegel, J. T., Navarro, M. A., Tan, C. N., & Hyde, M. K. (2014). Attitude-behavior consistency, the principle of compatibility, and organ donation: A classic innovation. *Health Psychology, 33*, 1084-1091.
- Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56*, 283-294.
- Stepanova, E. V., Strube, M. J., Clote, L. E., & Limes, D. (2016). Pictorial race activation in priming measures. *Basic and Applied Social Psychology, 38*, 223-239.
- Todd, A. R., & Simpson, A. J. (2017). Perspective taking and member-to-group generalization of implicit racial attitudes: The role of target prototypicality. *European Journal of Social Psychology, 47*, 105-112.
- Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science, 27*, 384-393.
- Veenstra, G. (2013). The gendered nature of discriminatory experiences by race, class, and sexuality: A comparison of intersectionality theory and the subordinate male target hypothesis. *Sex Roles, 68*, 646-659.
- Weigel, R. H., & Newman, L. S. (1976). Increasing attitude-behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology, 33*, 793-802.
- Weigel, R. H., Vernon, D. T., & Tognacci, L. N. (1974). Specificity of the attitude as a determinant of attitude-behavior congruence. *Journal of Personality and Social Psychology, 30*, 724-728.
- Wilson, J. P., Remedios, J. D., & Rule, N. O. (2017). Interactive effects of obvious and ambiguous social categories on perceptions of leadership: When double-minority status may be beneficial. *Personality and Social Psychology Bulletin, 43*, 888-900.
- Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods, 49*, 1193-1209.