

To appear in: J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.) (2021), *The Cambridge handbook of implicit bias and racism*. Cambridge, England: Cambridge University Press.

## REPORT FROM THE NSF CONFERENCE ON IMPLICIT BIAS

September 22, 2020

### Reporting Committee Members:

Bernadette Park (University of Colorado, co-chair)  
Richard E. Petty (Ohio State University, co-chair)

Joshua Correll, University of Colorado  
Lisa Feldman Barrett, Northeastern University  
Vince Hutchings, University of Michigan  
Gary Langer, Langer Research Associates  
Sabine Otten, University of Groningen  
Christopher Parker, University of Washington  
William von Hippel, University of Queensland

### Overview

Over the last several years, the study of implicit bias has taken the world by storm. Implicit bias was even mentioned by then candidate, Hillary Clinton, in a presidential debate in 2016. She went on to claim that implicit bias has deadly consequences when black men encounter law enforcement (for example, see Correll et al., 2007; Correll, Park, Judd, & Wittenbrink, 2002; Eberhardt, Goff, Purdie, & Davies, 2004). The controversy over police shootings of black men and women has only intensified as evidenced by public outcry over the murder of George Floyd on May 25, 2020, and increasing public support for the “Black Lives Matter” movement and its calls for liberty, justice, and freedom (Cohn & Quealy, 2020). These current events are but one reason why the study of implicit bias has so captivated the attention of the larger public: reducing it seems to have the potential to solve real-world problems. One idea is that if police officers were made aware of their implicit bias or participated in training workshops to reduce implicit bias, then perhaps fewer black people would wind up dead, arrested, or disproportionately sentenced to receive the death penalty (Baumgartner, Epp, & Love, 2014; Eberhardt, 2020).

Still, for all of its promise, contemporary scholars examining prejudice continue to struggle with the science of implicit bias. Most fundamentally, these struggles concern not only definitional issues (i.e., what is meant by implicit bias?) but also measurement issues such as how to best capture implicit bias. For example, researchers struggle with the reliability of implicit measures of attitudes, with the lack of correlations between alternate implicit measures of the same attitude, with low implicit to explicit attitude measure correlations (more so in some domains than others), and with

the ability of scores on implicit measures to predict discriminatory behavior (e.g., Schimmack, 2020). Although low correlations with explicit attitude measures are not in and of themselves a death knell (after all, implicit measures were developed precisely because it was thought they could capture something other than what explicit measures tap), the inability to predict or only weakly predict prejudicial behavior is more problematic. Although the link between attitudes and behavior has long been controversial (e.g., Ajzen & Fishbein, 1977; Wicker, 1969), part of the enthusiasm around the notion of implicit bias was the belief that it would permit researchers to bypass the tendency for people to edit their attitude reports on sensitive issues, tapping into what respondents “really think” about the attitude object. That is, if people wish to avoid being viewed as bigoted, they may be motivated to conceal their prejudice (e.g., Dunton & Fazio, 1997).

Implicit measures of attitudes, because they are less susceptible to strategic control, were believed by some to be more authentic than their explicit counterparts. But if their ability to predict behavior is limited, this raises concerns about the usefulness of such measures (for examples of their predictive utility, see Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Greenwald, Smith, Sriram, Bar-Anon & Nosek, 2009; Pasek, Tahk, Lelkes, Krosnick, Payne, Akhtar, & Tompson, 2009; Payne et al. 2010; Pérez, 2010; Ziegert & Hanges, 2005, and for evidence of their failure to predict, see Blanton et al., 2009; Kalmoe & Piston, 2013; Kinder & Ryan 2017).

Critically, in part because of the interest in changing prejudicial behavior, interventions based on current knowledge about implicit bias may be running ahead of the science. There is, at present, a push for anti-bias training (often conflated with implicit bias training) within government and private sector organizations (e.g., the Starbucks anti-bias training, Avila, Parkin, & Galoostian, 2019), when what we need is better *evidence* about the impact of implicit bias in society and how best to counter or diminish the likelihood of implicit bias operating. That is, we need an evidence base for making recommendations regarding whether and how implicit bias operates and how it can be reduced. Specifically, we need an evidence base for understanding how best to minimize its effects on decision and behavioral outcomes (see Dobbin & Kalev, 2018).

In this report we employ a multi-faceted definition of implicit bias and present relevant examples from the literature suggesting distinct types of implicit bias. We next present a schematic representation of attitudes, implicit and explicit measures of attitudes, and their link to behavior. One of the difficulties around the science of implicit bias is that different researchers use the same terms to mean very different things. The principal goal of the schematic is to make clear our terminology, as well as our working assumptions, so that readers of this report can have a clear understanding of these to help inform their own acceptance or rejection of the claims made here. Next, we provide a brief overview of the way that the field of social psychology has studied implicit bias. We then summarize some important points of agreement within the field regarding implicit bias and implicit measures of attitudes, as well as some unresolved issues.

Based on our understanding of where the science of implicit bias currently stands, we articulate important directions for research going forward. We end with some cautionary notes regarding the application of the construct of implicit bias in popular discourse and policy making, with the hope

of allowing the science to catch up to the enthusiasm for the ideas and concerns that have led to so much interest in implicit bias. In the final analysis, researchers, grant funders, and policy makers must not lose sight of the key end goal of measuring, documenting, and ultimately reducing implicit bias – to improve outcomes for people as they navigate their lives in the real world, both as behavioral agents and as members of marginalized groups.

## I. NSF Conference on Implicit Bias

On September 28-29, 2017, NSF convened a meeting to address the current state of knowledge regarding implicit bias as developed largely in social psychology. This report is organized around the key issues addressed at the meeting including: the definition of *implicit bias* and how it is similar and different from related concepts such as *implicit measures* and *implicit attitudes*, what is known about the phenomenon (i.e., the general consensus in the field), what is uncertain (i.e., divergent views in the field or insufficient research available), and what new research is recommended.

## II. Defining Implicit Measures, Attitudes, and Bias (as Compared to Explicit Measures, Attitudes, and Bias)

Before turning to a review of what is known and what remains to be studied, it is important to define what we mean by the major concepts and distinctions at use in the field of implicit bias. One key distinction is between implicit versus explicit *attitudes* and implicit versus explicit *measures* of attitudes. After defining these, we turn to the notion of implicit versus explicit bias.

**Implicit versus explicit measures.** An *explicit measure* is one that is transparent. It is clear to the person being assessed what is being measured. An *implicit measure* attempts to assess a belief, attitude, or behavior without the person's knowledge. For example, an explicit measure of attitudes toward black people might directly ask the respondent, "Do you like black people?" An implicit measure of attitudes would aim to assess liking without directly asking the person, such as by measuring how closely the respondent sits next to a black person. In this sense, contemporary implicit measures are in the same category as classic "indirect" measures as they require the assessor to make an inference about the construct of interest from some other response (Petty & Cacioppo, 1981). If we think of a continuum of implicit to explicit measures, this continuum would map onto the extent to which people were aware of the fact that the measure was attempting to assess their belief, attitude, or behavior. An *implicit measure* is one for which people have relatively low awareness of what is being assessed, whereas for an *explicit measure*, awareness is relatively high. A perfect implicit measure would assess the relevant construct without the person's awareness, and a perfect explicit measure would be one in which the person was fully aware. As we describe in some detail later, psychologists have developed a wide variety of implicit measures relevant to understanding bias. These measures have focused on assessing peoples' overall attitudes and stereotypes toward various minority groups.

**Implicit versus explicit attitudes.** Just as an explicit measure is a measure in which people are aware, an *explicit attitude* is an evaluation that people consciously and willfully acknowledge

(e.g., I like black people). In contrast, an *implicit attitude* is often defined as one that people hold but do not recognize or endorse (e.g., Kihlstrom, 2004).<sup>1</sup> Explicit measures, by definition, assess explicit attitudes. However, it is not the case that implicit measures necessarily assess implicit attitudes. That is, even if people are not aware of what a measure is attempting to measure (e.g., such as measuring liking by unobtrusively assessing how close one person sits to another), what is assessed with this measure could still be an explicit attitude – an attitude that the person would have reported if they had been asked directly (i.e., people may sit closer to people that they explicitly like). Indeed, in many domains, explicit and implicit measures of attitudes are correlated with each other, though the magnitude of the implicit-explicit relationship varies with the subject matter. For example, correlations are relatively high for political candidates but quite low in socially sensitive domains such as race (e.g., Greenwald, Smith, Sriram, Bar-Anon & Nosek, 2009).

The same factors that contribute to the link between implicit and explicit measures of attitudes being strong also contribute to the ability of implicit measures to predict behavior. For example, just as implicit measures on sensitive issues correlate less well with explicit measures than for non-sensitive issues, so too do they predict behavior less well in these domains. Indeed, meta-analyses show that the more strongly implicit measures of attitudes correlate with explicit measures, the more strongly implicit measures predict relevant behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Kurdi, et al., 2018). A situation of considerable interest in the field occurs when attitudes assessed with explicit measures and implicit measures fail to correspond (e.g., a person is positive on an explicit measure but negative on an implicit measure), a situation that has been called *implicit ambivalence* if people are unaware of the discrepancy (Petty, Tormala, Brinol, & Jarvis, 2006).

**Implicit versus explicit bias.** Having defined implicit versus explicit measures and attitudes, what about implicit versus explicit bias? Bias with respect to people occurs when an individual has more favorable beliefs, attitudes, and/or actions toward a person who is a member of a certain category (e.g., White, Female) than another category (e.g., Black, Male) in the absence of any individuating information about that person that would justify that favorability. *Explicit bias* occurs when people are aware of their bias and consciously and deliberately act on their acknowledged prejudicial attitude (e.g., I don't like African-Americans so I will recommend against hiring them). That is, people are aware of being biased and of acting on their bias. Importantly, if a person is fully aware of being biased in both evaluation and action, but the person simply aims to conceal this bias when asked about it, this person still has an explicit bias (i.e., has an explicit attitude and engages in explicit behavior). Though, due to social desirability concerns, it may be difficult to uncover this on explicit measures. That is, if people know that someone is attempting to assess their potentially biased attitudes or behaviors, they could deliberately conceal their responses. Indeed, getting around socially desirable responding was one reason that implicit measures of attitudes were developed in the first place (Fazio, Jackson, Dunton, & Williams, 1995).

---

<sup>1</sup> One could similarly apply this distinction to beliefs and behaviors such that explicit beliefs and behaviors are those that are willfully held or intentional, whereas implicit beliefs and behaviors are those that are held or occur unintentionally or out of awareness.

In contrast to a bias of which a person is aware, *implicit bias* also refers to the prejudicial judgments, decisions, and behaviors a person enacts, but there is something about the bias about which the person is unaware. Since the start of research on implicit bias in social psychology (Greenwald & Banaji, 1995), scholars in the area (and even the general public) have used the term implicit bias to refer to one of three things about which people may be unaware (see Petty, Wheeler, & Tormala, 2003). First, people can be unaware of the biased attitude itself. In this case, one can say that the person has an *implicit attitude*. Second, people can be unaware of the impact of a biased attitude on other judgments and behavior. In this case, one can say that there is an *implicit impact* of the biased attitude. Third, people can be unaware of the underlying source (basis) of their biased attitude. In this case, one can say that the attitude has an *implicit basis*. We provide an example of each kind of implicit bias next.

- (1) *Implicit attitude.* As noted earlier, a person may not be aware of having a reaction to one group that is more negative than another group. For example, people may not realize that they harbor more negative feelings toward black people than white people and would rate both groups equally on an explicit measure such as a direct self-report. When people are unaware of their attitudes, these attitudes are implicit. Yet, implicit (unaware) attitudes might still have an impact on behavior. However, there is relatively little evidence for the notion that people lack awareness of their attitudes. That is, there is little or no compelling research to date showing that people have no inkling of their evaluative responses. For example, a person reporting an explicit positive attitude toward a minority group might still recognize having an initial negative reaction that could be reported if the explicit measure had asked about “gut reactions” rather than an overall evaluation (Jordan, Whitfield & Zeigler-Hill, 2007; Ranganath, Smith, & Nosek, 2008). Or, people might be aware of an initial negative reaction but believe that this is invalid information because it reflects social stereotypes that they do not endorse, and thus do not report this on the typical explicit measure (Loersch, McCaslin, & Petty, 2011; Petty et al., 2006). What is more likely than people being completely unaware of their evaluative reactions is that they may not always appreciate that their automatic evaluative reactions can differ from their more deliberative evaluations and yet still have an impact on their behavior.
- (2) *Implicit impact.* In this definition, a person may be fully aware of a prejudicial negative reaction (whether automatic or deliberative) to a group member (i.e., the person has an explicit attitude), but the person may not be aware that this reaction is affecting his or her judgments or behavior (e.g., biasing perception of the person). This biasing impact can be completely unintended (e.g., see Greenwald & Banaji, 1995). Although people might be unaware of how various explicit attitudes affect their behavior, what people may be especially unaware of is how their relatively automatic reactions to other people can affect their behavior in addition to their more deliberative views. That is, as just noted, people may be aware of their tendency to have a quick negative reaction to someone (i.e., the reaction itself is not implicit), but they may believe that only their more carefully considered views affect their behavior and not their more automatic reactions. The fact that people can behave in a biased way without intending to be biased or “without even realizing it” (Kang, Bennett, Carbado & Casey, 2011) is perhaps the most common use of the term implicit bias.

(3) *Implicit basis.* Some theorists refer to implicit bias as including situations in which a person is aware of his or her attitude and is also aware of the effect it has (failing the first two definitions above), but the person is simply not aware of from where the attitude comes. For example, an attitude could be driven by underlying stereotypic expectations and the person does not realize this is the case (cf., Wilson, Lindsey, & Schooler, 2000). This would include misattribution effects where, for example, the evaluation is negative (I am aware I dislike Hillary Clinton as a presidential candidate and that as a result I will not vote for her), but the person doesn't realize that this negativity is a product of a group stereotype (e.g., a mismatch between one's expectation for what is "presidential" and what a typical woman is like). Moss-Racusin, Dovidio, Brescoll, Graham, and Handelsman (2012) show that science faculty judge a female applicant to be less competent and less hireable than an identical male applicant for a position managing a science laboratory, presumably not because they believe females to be incapable of succeeding in science (the effect was equally strong among female as male science faculty), but rather because the application when paired with a female name did not evoke the same sense of competence and hireability. In this version of implicit bias, the person fails to know the stereotypic basis of the negative attitude.

Accepting these three options implies that unawareness of the attitude is not the only or even a necessary characteristic of implicit bias since as explained, a person can accept that their attitude is biased but not realize that this attitude influences them or what the basis of the attitude is. In short, the overarching characteristic of all three forms of implicit bias is that the phenomenon involves something of which the respective person is unaware.<sup>2</sup>

These issues regarding lack of awareness of 1) the attitude, 2) its impact on behavior, and 3) its origins can raise complex issues with regard to *personal responsibility*. For example, if a person is completely unaware that a biased attitude is affecting his or her behavior, should the person be held accountable for this behavior? Notions of accountability in most societies are predicated on the belief that individuals are responsible for their own beliefs and actions, but numerous models in social psychology raise questions about this bedrock belief. For example, some scholars have proposed that society is so infused with stereotypes that they are activated automatically whenever group members are encountered. What differentiates prejudiced from non-prejudiced people in this model is whether the automatically activated stereotypes are subsequently used to guide judgment and behavior or are inhibited (Devine, 1989). One implication of this notion is that people can no more prevent activation of their stereotypes than they can prevent themselves from reading a stop sign once they have seen it (for early refinements of this model, see Gilbert & Hixon, 1991; Spencer, Fein, Wolfe, Fong, & Duinn, 1998). Subsequent research took questions of accountability

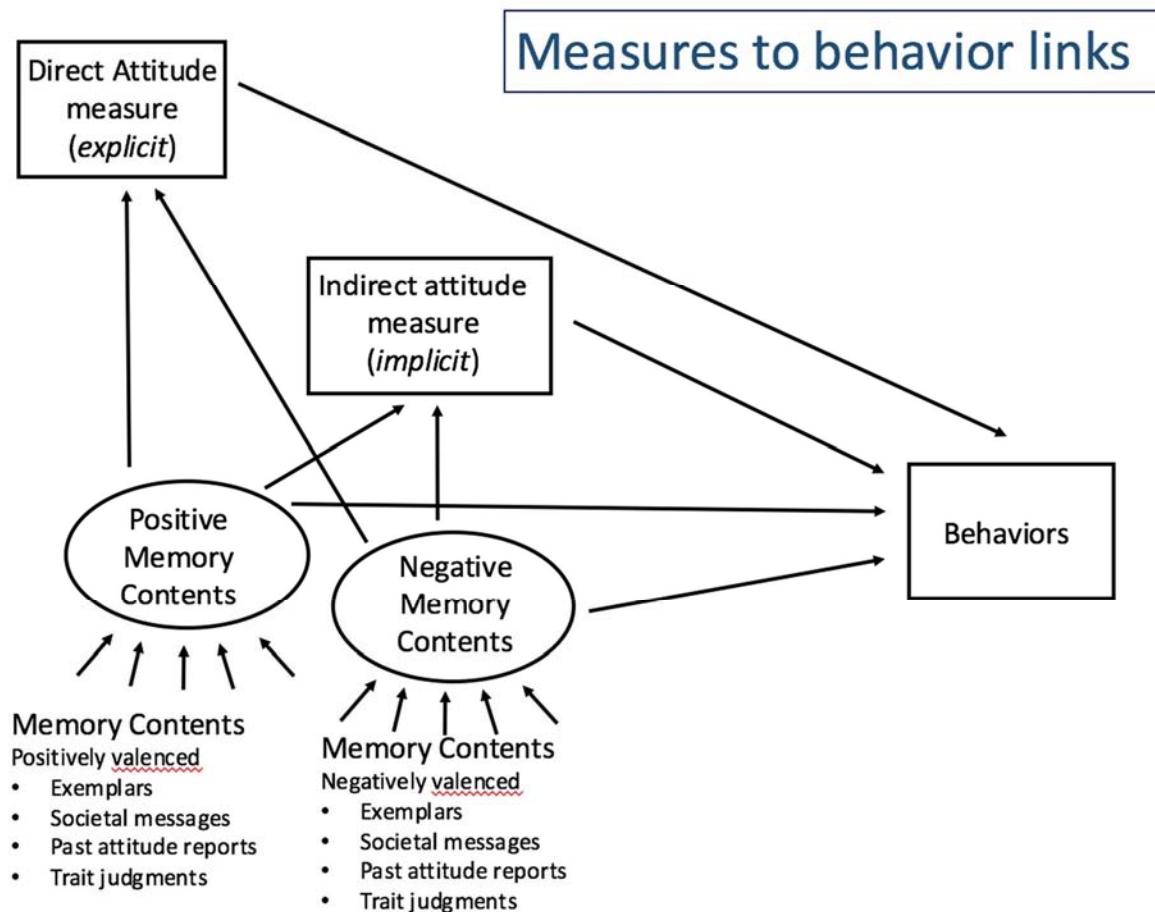
---

<sup>2</sup> Although we identify the term implicit with lack of awareness (cf., Kihlstrom, 2004), some researchers argue that "the term implicit can best be understood as being synonymous with the term automatic" (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). If so, one still would need to distinguish automatic from controlled *attitudes*, an automatic from a controlled *impact* of an attitude, and an automatic versus a controlled *basis* of the attitude. Finally, some scholars equate studies of implicit bias with the use of implicit measures (Greenwald & Lai, 2020). However, when implicit and explicit measures of attitudes correlate highly and predict the same outcomes, this definition makes it difficult to distinguish implicit from explicit bias effects.

one step further by documenting that older adults have difficulty inhibiting their automatically activated stereotypes due to compromised frontal lobe functioning (Stewart, von Hippel, & Radvansky, 2009; von Hippel, Silver, & Lynch, 2000). This finding raises the possibility that sometimes neither activation nor application of stereotypes can be prevented, suggesting that it is unclear whether people should be held personally responsible for their implicit biases.

### III. A Schematic Representation of the Current Conceptualization of Implicit and Explicit Measures of Attitudes

In an effort to be clear about the constructs we are describing, the following schematic is offered. It is rough, incomplete, and many at the conference and in the field more generally will likely disagree with some part of it. But some of the controversy regarding the body of research on implicit measures, attitudes, and biases comes from having different conceptions of what implicit versus explicit measures are assessing and especially the relationship between measures and behaviors. Our goal is to at least be clear about some definitional issues to help remove disagreements that stem from a basic misunderstanding or confusion about what researchers are talking about.



Next, we provide a few important points regarding the schematic and its implications for understanding implicit measures and bias.

- (1) As depicted in the figure, attitude measures, whether implicit or explicit, tap into the same contents in memory. This information can be evaluatively positive or negative and these two sources of evaluative information can influence both explicit and implicit measures of attitudes (that is, this need not be a bipolar evaluation, see Cacioppo & Berntson, 1994). The memory contents can include attributes of the attitude object, specific past encounters with the attitude object, emotions, societal and media messages, past attitude reports, etc.
- (2) As implied by the figure, an attitude (or evaluative reaction) is not necessarily a “thing” that exists in memory as a unified entity and retrieved automatically and inevitably, but rather can be at least partially constructed “on the fly” in response to a particular attitude object in a given situation or in response to an explicit question. Furthermore, evaluations, such as attitudes toward minority groups, can vary with the context such as the race of the experimenter taking the measurement (Lowery, Hardin, & Sinclair, 2001) or the clothing worn by the target to be judged (Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Judd, & Park, 2001). Thus, it is not surprising that the measured attitude toward various objects as assessed with both implicit and explicit measures can vary somewhat from situation to situation.
- (3) Although as just noted, the implicit and explicit measures are informed by the same underlying memory contents, the two kinds of measures are not invariably related to each other and will not necessarily yield the same evaluative outcomes (i.e., the same attitudes) for a variety of reasons:
  - a. Moderators (e.g., motivation and opportunity to think) affect both kinds of measures and can do so to different degrees.
  - b. Different contexts or environments can make different memory contents more or less accessible automatically and deliberatively.
  - c. To the extent that the memory contents activated through automatic processes differ from the contents activated by controlled processes, coupled with the fact that some implicit measures are more influenced by automatic processes and less influenced by controlled processes relative to explicit measures, this can produce differences on the implicit versus explicit measures.
  - d. Automatic and controlled processes may access or prioritize different things such as positive versus negative memory contents or affectively versus cognitively based contents.
  - e. Often researchers are asking different questions with an implicit versus explicit task, and this contributes to the divergence of outcomes (for demonstrations of this, see Han, Czellar, Olson, & Fazio, 2010; Payne, Burkley, & Stokes, 2008; Wittenbrink et al., 2001).

- (4) Contemporary implicit measures primarily utilize automatic processes, whereas explicit measures are more influenced by controlled processes (Schneider & Shiffrin, 1977), but neither measure is process pure (e.g., Calanchini & Sherman, 2013; Calanchini, Lai, & Klauer, 2020). Thus, although implicit measures are heavily affected by automatic processes and less so by controlled, and explicit measures are driven more by controlled than automatic processes, the processes themselves are separate from and not tied in an isomorphic manner to the measures.
- (6) The act of completing an implicit measure can influence responses on an explicit measure (e.g., one might respond differently to a thermometer rating judgment having just completed an Implicit Association Task (IAT)). Similarly, completing an explicit measure might affect responses on an implicit measure by making certain memory contents more salient.
- (7) The link between attitudes and behavior has a long history in social psychology and this linkage has been shown to be dependent on a variety of moderators such as the correspondence of measurement of attitude and behavior (Fishbein & Ajzen, 1977), the strength of the attitude (e.g., its certainty, accessibility; Petty & Krosnick, 1995), and other factors. Because of established moderators, neither explicit nor implicit measures of attitudes will necessarily predict behavior in all circumstances. Furthermore, sometimes, one kind of measure will predict behavior better than the other kind. For example, prediction of each kind of measure tends to be better the more the behavioral criteria match the attitude measure – in deliberative situations, explicit measures tend to predict better, but in spontaneous situations, implicit measures tend to predict better (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). Although the field's understanding of moderators of the ability of explicit attitudes to predict behavior is fairly well developed, the same is not true with respect to implicit measures.

#### **IV. Implicit Measures that Have Been Developed to Study Implicit Bias**

As noted earlier, social psychologists have developed implicit measures that are designed to assess people's attitudes without asking them directly. When such measures were first introduced many decades ago, they were called *indirect measures* because a person's attitudes needed to be inferred indirectly, typically from some behavioral response rather than from a direct self-report (Petty & Cacioppo, 1981). For example, rather than directly asking Person A how much they like Person B, liking could be inferred from spontaneous eye gaze or seating distance (Dovidio, Kawakami, & Beach, 2001). Or, rather than asking someone how much they liked religion, they could be asked to tell a story about a minister at the pulpit and the story be coded for positive versus negative content (Proshansky, 1943).

In the past few decades, a new type of implicit measure was developed. What set these measures apart was that attitudes were inferred from how quickly people could give evaluative reactions to various stimuli. That is, these measures attempted to assess automatic evaluative reactions. Although it could be argued that the earlier indirect measures, such as seating distance and

telling spontaneous stories, were successful because they also tapped into automatic reactions (e.g., people spontaneously sit closer to others they like than dislike without giving it much thought), the new implicit measures were quite explicit about the desire to assess automatically activated attitudes. These measures include the Implicit Association Test (IAT; Greenwald, McGhee & Schwarz, 1998), the evaluative priming measure (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), and the Affect Misattribution Procedure (AMP; Payne & Lundberg, 2014), among others (for reviews, see Petty, Fazio, & Briñol, 2009; Wittenbrink & Schwarz, 2007).

In their initial development during the 1980s, a primary goal of the new wave of implicit measures, like their predecessors, was to minimize strategic responding with respect to sensitive or socially undesirable attitudes as it was assumed that quick responding would likely bypass social desirability concerns (e.g., Dovidio, Evans, & Tyler, 1986; Gaertner & McLaughlin, 1983). Logically, there are multiple ways to achieve this goal (e.g., the bogus pipeline method; Jones & Sigall, 1971). The newer reaction time-based implicit measures draw on paradigms from cognitive psychology that examine facilitation and inhibition effects in priming (Meyer & Schvaneveldt, 1971; Stroop, 1935). For example, participants' attitudes about race are inferred by assessing the degree to which race-related cues (e.g., pictures of black and white faces) interfere with or facilitate some non-racial judgment (e.g., categorizing a word as positive or negative; Fazio, Jackson, Dunton, & Williams, 1995).

By way of example, an implicit association test (IAT; Greenwald et al., 1998) examining racial attitudes might ask participants to alternate between classifying a set of valenced words as either good or bad (a task that is unrelated to race) and classifying a set of names (e.g., Jamal, George) as either stereotypically Black or White (a task presumably unrelated to attitudes). By forcing the participant to use the same keys for both tasks and by flipping the mapping of the keys for one categorization task in the middle of the larger task, the researcher can assess the degree to which classifying a word as good is easier when it shares a response key with White and harder when it shares a response key with Black. Similarly, an evaluative priming task (Fazio et al., 1995) might present a series of words on a computer screen and ask participants to classify each as either good or bad. This classification task has nothing to do with race – words (e.g., happy) are simply categorized as good or bad. Immediately before each word, however, a prime appears – perhaps a face that is either a black person or a white person. Bias is measured as the extent to which the race of the face interferes with or facilitates the judgment of the word as good or bad. Different measures adopt slightly different approaches (some rely on response competition, others on semantic priming, etc.), but almost all of the measures use some variation of this indirect strategy.

Most tasks also share a number of interesting, but likely unnecessary, characteristics. Perhaps due to their structure, perhaps due to the fact that these approaches were borrowed from cognitive psychology, most measures typically involve indices derived from reaction times or error rates. Accordingly, participants must generally perform the tasks on computers, ideally in a distraction-free environment. And most measures involve some degree of difficulty, presumably because if the focal task is too easy, responses may be impervious and therefore insensitive to race or gender or whatever dimension the researcher hopes to assess. Some implicit measures (such as the Go/No-go association test, or GNAT, Nosek & Banaji, 2001) impose a strict response deadline, forcing the participant to respond quickly and inducing errors. Other tasks,

such as a first-person-shooter task (Correll et al., 2002), present ambiguous stimuli, forcing the participant to interpret the information (e.g., is the target holding a gun or a cell phone) and increasing mean response times. Still other tasks, like the IAT, force the participant to perform cognitively demanding mental operations as they alternate between judgments.

In considering new approaches to implicit (indirect) measurement, it is worth revisiting the original goal of this class of measures. There are several reasons for doing so. First, it is important to consider whether or not the original goal (removing social desirability) is still relevant. Based on what researchers have learned over the past 30 years, there may be new considerations that should guide the field (or that may guide an individual researcher). And, for whatever new goals may be deemed relevant, scholars might consider whether the approach that has generally been adopted (difficult indirect measures that rely on response times or error rates) offers the best strategy for achieving those goals. For example, if the goal of contemporary implicit measurement is to get at gut reactions (rather than avoid social desirability), explicit measures that ask about gut feelings may accomplish a similar purpose (e.g., Ranganath et al., 2008).

**Potential goals of implicit measures.** This discussion will only address a few of the potential goals of the currently popular implicit measures. First, it is noteworthy that despite social desirability concerns, people's explicit attitudes can still show evidence of bias. For example, Piston (2010) found that 45% of white people interviewed in the face-to-face portion of the 2008 National Election survey rated Blacks lower than their own racial group on the trait of "hard working." The figure was 39% on the trait of "intelligence." Nonetheless, because many people may be uncomfortable expressing racist or sexist attitudes in a public survey, measures that reduce socially desirable responding still offer value.

In addition to social desirability, there are other considerations that the field has come to recognize more fully over the last few decades of research. One, already discussed, is that it may be desirable to have measures that assess automatic as well as deliberative attitudes and that the new implicit measures focus on the former. This is a value above and beyond avoiding social desirability. Another factor is that participants may not have complete introspective access to all of the mental content that is associated with an attitude object. For example, as noted earlier, although depictions of different racial groups in the media might produce negative associations to racial categories, people do not consider this when they report their explicit attitudes. Thus, another goal for implicit measures could be to assess these "extra-personal" associations that exist in the mind of an individual because they exist in the culture in which that individual lives (Olson & Fazio, 2004). The importance of cultural associations on what is assessed by implicit measures is evident in data showing that IAT responses are associated with a person's geographic location (Hehman, Flake, & Calanchini, 2018; Payne, Vuletich, & Lundberg, 2017). Even in domains that are not particularly sensitive to social desirability concerns, participants may not actually be able to report all of the content that underlies their attitudes (i.e., as noted earlier, some attitudes can have an implicit basis). One of the virtues of the current class of implicit measures is that, because they never directly ask the question, they do not require that participants consciously access (or construct) the relevant attitude.

Once the field's goals for implicit measures have been clarified, the particular characteristics of those measures should be revisited. Again, the vast majority of them involve indirect assessment. It should be clear that, if researchers choose to assess attitudes about race or gender in the context of some other task (e.g., classifying words as good or bad), the ability to assess the attitude in question is necessarily constrained to some extent by the features of that task. For example, the evaluative priming task described above requires a set of cognitive operations (perceiving the text, interpreting it, and classifying it according to valence) that are distinct from the attitude being measured but that may operate as a kind of filter for the associations that form that attitude. Is this approach necessary? Is it optimal? Are there other ways to achieve the goals of these measures without introducing an unrelated task? These tasks also generally involve some non-trivial level of difficulty. Again, does this approach serve the goals of these measures? These are important questions to ponder. In some cases, it may be that carefully designed and carefully administered explicit measures meet research goals as well or better than implicit measures. Indeed, while the present committee's work focuses on the state of the science on implicit bias, further study into the optimal measurement of explicit bias is important to pursue as well.

**Atypical measures.** Although the previous discussion has focused on some of the most popular implicit measures currently in use, there are a variety of other measures that deviate, in one way or another, from the typical cognitively-inspired structure that characterizes measures like the IAT, GNAT, evaluative priming, and the first-person-shooter task. These atypical paradigms may offer hints and inspiration that can facilitate development of new, different, and better measures. For example, consider the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun & Stewart, 2005). Like the other measures, the AMP relies on indirect measurement, but it does not require measurement of response time or error rates. Drawing on Murphy and Zajonc (1993), the AMP presents a prime stimulus (e.g., a white or black face) followed by a Chinese ideograph. Participants, who typically do not read Chinese, are asked to evaluate the complex but (to them) meaningless ideograph. Those evaluations are typically biased by the prime. That is, a prime about which a person feels positively biases responses to the Chinese ideograph in a positive way.

A second atypical example is one of the few tasks that does not rely on indirect measurement. This is actually an explicit measure in that it directly asks the participant to evaluate or classify the attitude object that is being studied. This measure relies on a mouse-tracking paradigm (Freeman & Ambady, 2010), which has been used to study how people classify ambiguous stimuli. The paradigm presents two response labels (e.g., "good" and "bad," "male" and "female," or "caring" and "aggressive") at the upper left and upper right corners of a computer screen. Participants must use a computer mouse to position the cursor at the bottom center of the screen. They are then shown an exemplar and asked to move the cursor to the appropriate response option. This work showed that, when presented with prototypic exemplars (e.g., a very masculine-looking male face), participants moved the cursor toward the appropriate option (e.g., "male") in a relatively straight line. When presented with a less prototypic exemplar (e.g., a more feminine-looking male face), the movement was less direct: initially, the cursor might head toward the incorrect label ("female") before veering off toward the correct response. This approach has also been applied to the measurement of attitudes (Vallacher, Nowak, Froehlich, & Rockloff, 2002). For example, when asked to classify an attitude object like *sunshine*, trajectories

are relatively direct, but when asked to classify an attitude object like *euthanasia*, about which people may be more ambivalent, the mouse trajectories are less direct (Schneider et al., 2015; Vallacher, Nowak, & Kaufman, 1994). Thus, this task can assess ambivalence with respect to a target without ever asking about ambivalence.

Third, the recently proposed Judgment Bias Task (JBT; Axt, Nguyen, & Nosek, 2018) is an indirect measure that relies on much more deliberative processing than most of the other measures discussed here. Drawing on Beckett and Park (1995), this paradigm simulates a hiring or other selection task. It presents a series of applicant profiles, which the participant must evaluate. These profiles might include information about each applicant's grades, test scores, recommendation letters, etc. The participant then must decide whether or not to hire (or admit, or date) each applicant. Critically, although some applicants are more qualified and some are less qualified, the differences are hard to detect. And, just as critically, the profiles also include information that interferes with the judgment task, such as photographs that vary in attractiveness, or information that the applicant is an ingroup or outgroup member. The goal of the measure is to see how the variable of interest (e.g., gender) affects some outcome of interest (e.g., a hiring decision). For example, how well can participants distinguish objectively more from less qualified female applicants for a job and how lenient or strict are they in making these decisions.

Finally, there is growing interest in using brain imaging in the study of implicit bias. The neurons in a number of brain regions routinely alter their activity during brain imaging studies of implicit racial attitudes (e.g., Phelps et al., 2000). Neuroscientists tend to refer to brain regions in terms of their own interests and preferences, so it is not uncommon for neuroscientists who study implicitly measured attitudes and prejudice to refer to these brain regions as the “prejudice network” (e.g., Amodio, 2014). Unfortunately, this can lure consumers of neuroscience research to the misperception that these regions are consistently and specifically associated with implicit racial bias. To date, brain imaging research does not yet clarify the debate regarding the power of implicit measures of racial attitudes to impact prejudicial behavior. The regions and activity patterns associated with implicit measures of prejudice contribute to many different psychological events, and therefore with current technology, brain images are only useful correlates of implicit measures of bias when they are combined with behavioral measures. Even then, however, brain imaging findings do not transparently explain the mechanisms behind what implicit measures assess. People might attend to other-race faces because those faces are novel. Similarly, people might better remember faces of the same race as a function of greater expertise. In fact, the novelty and lack of perceptual expertise (associated, as they are, with increases in arousal) may be a component of implicitly measured attitudes. Thus, linking implicit bias with brain imaging is a challenging task, but one that may ultimately yield important insights depending on the specific goals for such measures.

In sum, there are a variety of implicit measures of attitudes that have emerged, each with idiosyncrasies and each with particular strengths and weaknesses. The goal here is not to provide a catalogue but merely to highlight a few of the most commonly used measures, as well as those that deviate from what has (for better or worse) become the standard set. Future measures need not rely on indirect assessment, they need not rely on speeded decision making, and they need not involve response times and error rates. Perhaps by thinking beyond the default approach,

researchers can build measures that more effectively meet the desired goals whether those goals are to: (1) avoid social desirability, (2) assess automatically activated attitudes, (3) bypass participants' understanding of the basis of their attitudes, (4) tap evaluations of which someone is unaware, or some other goal.

## V. What Does the Field Agree Upon Regarding Implicit Bias and Implicit Attitude Measures?

A number of points seem to generate agreement among researchers regarding the study of implicit bias and implicit attitude measures.

- (1) To optimally compare the impact of explicit (self-report) and implicit (e.g., IAT) evaluations, researchers should compare them at the same level of categorization (e.g., see Fishbein & Ajzen, 1974). For example, if an implicit measure assessed attitudes toward black people (good or bad associations), the explicit measure should not ask something more specific such as attitudes toward affirmative action policies. Similarly, if an implicit measure asks people to categorize specific exemplars of a group (e.g., pictures of famous powerful women), the explicit measure should not ask about the group at the general category level (e.g., judgments of what “women” are like).
- (2) Some new evidence has emerged demonstrating a relationship between community-level implicit associations and demographic characteristics and outcomes of those geographic areas (e.g., Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016; Payne et al., 2017). For example, one study showed that U.S. counties with higher levels of county-wide implicit racial bias also had greater racial gaps in infant health outcomes, even after controlling for relevant demographic and geographical factors (Orchard & Price, 2017). This suggests that responses on implicit measures aggregated across individuals reflect behavioral characteristics of those communities, even though the correspondence of individual-level implicit measures to behavior has been relatively weak. At least in part this is likely because the community level measures aggregate across large numbers of observations and hence are much less “noisy” or unstable. A reasonable next step would be research examining the ability of measures of an individual aggregated across many contexts and situations to show enhanced test-retest reliability and better ability to predict that individual’s behavior (also across situations). Indeed, recent research is showing (in accord with psychometric theory) that when a person takes multiple versions of the same IAT across two different time periods and these assessments are averaged, test-retest reliability is increased compared to a single administration at each time period (Connor & Evers, 2020; Lindgren et al., 2018).
- (3) As just noted, implicit measures of attitudes may be more useful for tapping into group-level automatic evaluations than individual-level evaluations because, as currently developed, the typical single administrations have too much unreliability (and contextual impact) to be highly useful in assessing the automatic component of attitudes for particular people (unlike explicit attitude measures). Thus, if group A scores higher than group B in prejudice on an implicit measure, it is reasonable to predict that group A will

demonstrate more prejudicial behavior than group B. It is less reasonable, however, to use the implicit measure to predict which particular members of group A will engage in prejudicial behavior. In this sense, implicit measures can be compared to imperfect disease predictors. Imagine a cancer diagnostic test that shows high scorers on the test are more likely to develop cancer overall than low scorers. But the screening test is imperfect so that of every 10 who score high on the test, 4 (on average) will develop cancer but of every 10 who score low, only 2 (on average) will develop cancer. Such a test could be useful for screening groups and making conclusions such as those who score high are two times more likely to develop cancer than those who score low. But the test is not useful in saying which 4 of the 10 who score high will get cancer and which 6 will not. The same is true for implicit measures of prejudice as they currently stand. That is, they can be useful in predicting that certain groups of people (e.g., high scorers) are more likely to engage in prejudicial behavior, but they are not as useful for determining which particular individuals among the high scorers will engage in prejudicial behavior.

- (4) As articulated above, there are numerous moderators of the impact of both implicitly measured and explicitly measured evaluations on behavior.
- (5) Activation of an attitude should be distinguished from application of that attitude. That is, in any given situation, an evaluation can automatically come to mind, but that evaluation need not drive behavior (e.g., if situational pressures are more salient, or if deliberative thought processes override the evaluation that comes to mind).

## **VI. What Issues Are in Some Dispute with Evidence or Adherents on Both Sides and thus Point to Clear Research Questions to Resolve? (i.e., These Issues Have Achieved Some Research Attention, but Are Far from Being Fully Resolved and thus Are Worthy Of Additional Research)**

- (1) It is not clear how much contemporary implicit measures of attitudes assess transient momentary states versus relatively stable traits or what combination of each.
- (2) What is the internal reliability of various implicit measures and test—retest reliability, and how does this vary across different attitude objects, situations, and people (e.g., those more likely to rely on intuition, see Pacini & Epstein, 1999)? In addition to aggregating over multiple assessments as noted earlier, how can reliability be increased for current measures? Can new measures be developed that are more reliable?
- (3) How highly related are various implicit measures to each other? Although methodological factors (e.g., using the same stimuli) can improve relationships, what else can improve consistency across measures? What level of consistency should be expected and what does lack of consistency imply (e.g., measurement error; measures are highly sensitive to the immediate context)?
- (4) How well do implicit measures correlate with explicit measures for different attitude objects and issues? As noted earlier, it is clear from meta-analyses that these correlations

vary across attitude domains. They seem to be higher for political attitudes and mundane attitude objects (e.g., consumer products), and lower in socially sensitive (e.g., racial) domains. They are lower still for attitudes towards the self (Bar-Anan & Nosek, 2014). It is not entirely clear what is the underlying moderator. Is it the social sensitivity of the attitude object? Is it the complexity of the knowledge structure regarding the attitude object (with greater complexity lending itself to greater contextual variation in the activated attitude)? What are the primary causes and consequences of implicit-explicit attitude discrepancies?

(5) How well do implicit measures predict behavioral outcomes alone, and over and above explicit measures? In what domains are implicit measures a useful supplement to explicit measures? In the domains in which they provide added predictive power, why does this occur?

(6) Manipulations that alter implicitly measured associations don't necessarily show parallel changes in behavior. This is not particularly surprising because manipulations that alter explicitly measured attitudes don't always alter behavior and there is much research on the conditions under which this is more or less likely to occur (e.g., see research on attitude strength; Petty & Krosnick, 1995). Three key questions for changing implicitly measured attitudes are: (1) When do such shifts also have downstream behavioral consequences? (2) How can the downstream consequences of such shifts be magnified? (3) When are changes in implicitly measured attitudes more impactful on behavior than changes in explicitly measured attitudes?

## **VII. What Areas of Research Should Be Encouraged Going Forward? (These Suggestions Stem from the Knowledge That Has Already Been Accumulated and the Issues Noted Above Where the Existing Research Has Not Provided Sufficient Clarity)**

### **A. Measurement Issues**

(1) A systematic assessment of the measurement properties (e.g., reliability, convergent, and predictive validity) of implicit measures of attitudes should be undertaken. Since its introduction more than 20 years ago, the IAT has more and more dominated the field and mostly replaced other paradigms to detect implicit biases. In reviews on the psychometric qualities of different implicit measures, the IAT and AMP score best (e.g., Bar-Anan & Nosek, 2014). Yet, the appropriateness and reliability of different implicit measures will largely depend on the content of the to-be-detected bias and the particular procedures used. For example, when trying to investigate the impact of race on police officers' propensity to use their weapon, using the shooter paradigm (e.g., Correll et al., 2002) obviously has a higher ecological validity than using a race-IAT. But how strongly will these two measures correlate?

(2) There are now many different types of implicit attitude measures, and it seems highly likely that different measures have different strengths and weaknesses. It also seems highly likely that different measures are more or less suitable for different research questions and populations. To date, the field is still lacking in theoretical and empirical

knowledge to better understand when and why different implicit measures are dissociated, and, in turn, which specific measures are best suited for specific aims, contexts, psychological processes, and behavioral predictions (see Brownstein, Madva, Gawronski, 2020). In other words, a better understanding of similarities and differences among various psychometrically sound measures of implicit attitudes will help the field generate more straightforward suggestions on which measure may best fit which research question. A greater understanding of the current crop of implicit attitude measures would allow researchers to be more thoughtful and systematic in their choice of measures when initiating data collection.

(3) Most measures of implicit bias rely on reaction time, but as explained earlier, not all of them (e.g., Affect Misattribution Procedure, Payne, Cheng, Govorun, & Stewart, 2005; Partially Structured Attitude Measures, Vargas, von Hippel, & Petty, 2004; Stereotypic Explanatory Bias, Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003; Linguistic Intergroup Bias, von Hippel, Sekaquaptewa, & Vargas, 1997). Greater attention should be given to the development of new measures that rely on properties of implicit measures of attitudes beyond strength of association. There has been almost no systematic study of the varying utility of existing measures nor have there been systematic efforts to develop new measures with known properties. An effort to tie the development of new measures to the different types of implicit bias (outlined at the beginning of this document) would be potentially valuable.

(4) Future research should explore the conditions under which implicit measures predict (a) explicit measures of attitudes, (b) outcome measures and behaviors, and (c) outcome measures and behaviors over and above explicit measures. Moreover, to what extent do implicit and explicit attitude measures affect one another? Do implicitly and explicitly measured attitudes interact in causing behavior or do they exert orthogonal effects (e.g., Johnson, Petty, Briñol, & See, 2017)?

## B. Questions of Causation

(5) What role do implicit attitudes and implicit bias more generally play in *causing* prejudicial behavior? To date, there has not been enough focus on studies of behavioral outcomes. Such studies should be prioritized due to their applied importance; the idea of implicit bias has captured the public imagination largely because of its presumed potential to explain discriminatory behavior outside the laboratory. For example, there are studies that reveal relationships between implicitly measured attitudes and important discriminatory behaviors outside the laboratory (e.g., Hehman et al., 2018), however, the causal role for the attitudes that implicit measures assess is unclear. Experimental work is necessary to establish the role of such attitudes in causing discriminatory behavior, but there are precious few laboratory and field studies that rely on experimental manipulation and measure discriminatory behavior. It will be particularly important for such experimental work to disentangle the influence of implicitly and explicitly measured attitudes, thereby establishing the unique predictive role of implicit measures in accounting for discrimination. It remains unclear whether the attitudes assessed with implicit measures are responsible for a) individual differences in how members of

outgroups are treated, and/or b) differential outcomes that emerge as a function of group membership (e.g., different job or school performance among people of different genders, ethnicities, etc.).

(6) Can interventions targeting implicit bias *reduce* prejudicial behavior? As noted earlier in our definition of implicit bias, there are several aspects of implicit bias of which people can be made aware. For example, people can be made aware of the fact that they hold prejudicial automatically activated attitudes, or they can be made aware of the fact that these automatic attitudes can influence their behavior without their awareness or intention. Some researchers claim that awareness of holding prejudicial automatic attitudes is helpful in reducing prejudicial behavior. Yet, this has not been examined carefully yet. Research is needed on the effectiveness of implicit bias workshops to better understand what psychological ingredients they contain, whether and under what conditions they work, and whether they could potentially produce boomerang effects. As governmental agencies and business organizations move to mandate implicit bias training, it is critical to establish a solid evidence base from which such interventions can draw. To what extent can scores on implicit or explicit attitude measures be changed by these interventions and what do such changes reflect? Is it driven by control or motivational processes in response to being made aware of biases? Is it necessary to change the underlying contents of the memory associations in order to reduce implicit bias, and is it reasonable that such trainings can do this? Or is a better strategy to educate perceivers about the conditions under which bias is most likely to operate (e.g., under high cognitive load, with ambiguous stimulus information), and to encourage them to monitor when such conditions are present in order to “check” on whether biases might be operating. For example, one technique is to ask perceivers to mentally simulate their response, changing the group characteristics of a target. If the Starbucks manager had asked herself, “would I call the police if these men were white?”, could this have helped her to identify the possible operation of implicit bias, and perhaps shift her behavior even if it didn’t fundamentally change the implicit bias itself?

### C. Animus Versus Tacit Acceptance of Existing Inequalities

(7) Although the literature on implicit bias understandably focuses on identifying, and ultimately reducing, biased outcomes, it is reasonable that the research would include as well actions that reflect tacit acceptance of existing racial (or other group based) inequities. That is, the study of implicit bias need not be confined to actions that are motivated by animus, unconscious, or otherwise. The robust literature on racial prejudice in political science and sociology identifying “new racism” theories such as Symbolic Racism (Kinder & Sears 1981; Sears, 1998; Sears & Henry, 2003; Sears & Kinder 1971; Tesler & Sears, 2010), Racial Resentment (Kinder, Sanders & Sanders, 1996), or Modern Racism (McConahay, 1983), maintain that contemporary racial bias relies less on overt claims of racial inferiority and is instead borne out of the belief that racial inequality is due to the cultural inadequacies of black people and other minorities. According to these theories, opposition on the part of white people to policies designed to address lingering racial inequity is derived from the widespread belief that African Americans and other minorities no longer face significant racial barriers following the passage of landmark

civil rights legislation in the 1960s. Bobo, Kluegel, and Smith (1997) adopt a similar perspective with their theory of Laissez Faire Racism. In brief, these authors argue that decades of overtly discriminatory policies during the period of Jim Crow have resulted in substantial racial disadvantage for African Americans in employment, educational attainment, household income, housing, and perhaps especially wealth. For example, they report that the average Black American family has only about one-tenth of the wealth of the average White American family. More recent studies confirm the persistence of this racial wealth gap for both Black and Latino families relative to White households (Darity et al., 2018). As with the other new racism theories, one of the pillars of Laissez Faire Racism theory is the denial of significant racial barriers in contemporary American society. Consequently, Bobo and his colleagues (1997) maintain that, “even if all direct racial bias disappeared, African Americans would be disadvantaged due to the cumulative and multidimensional nature of historic racial oppression in the U.S.” (Bobo et al., pg. 4).

Importantly, then, biased outcomes can be a product of the belief that historically disadvantaged groups now operate on a leveled playing field, and thus any existing inequalities must be a product of shortcomings of individual members of those groups. Therefore, the question arises whether in contexts with widespread and persistent racial inequality, indifference to structural hierarchies can be as consequential as blatantly (explicit) racist beliefs and/or more subtle implicitly measured attitudes. While this perspective may be present in current research efforts, it has not been directly targeted as a factor that could underlie implicit bias, and going forward it would be a fruitful avenue of investigation.

#### **D. Extensions of Existing Research**

(8) Effects of implicitly measured attitudes are sometimes stronger in the field than in the lab. It is not clear why this is the case. Is it strong situations versus weak situations? Is it the potency of the independent variables in the field versus the lab? Is it something else?

(9) NSF ideally can incentivize the field to do things that it won't or can't do on its own such as funding longitudinal, field studies. NSF should also consider funding consortium type research to pit different interpretations of important and robust effects against each other. As part of such an initiative, it would be important to establish a repository for all research, including and perhaps especially failed studies that were never published because they failed to find an effect. This information would helpfully inform researchers who begin a new and related research project and hopefully move the field along in a more efficient manner.

(10) A substantial portion of the literature on implicit bias is based on convenience samples. Such samples play a critical role in the development stages of research. Still, results from these samples cannot be assumed to generalize to the larger population, especially in the size of effects obtained (see, e.g., Callegaro, Villar, Yeager & Krosnick, 2014; MacInnis, Krosnick, Ho & Cho, 2019; Malhotra & Krosnick, 2007; Pasek & Krosnick, 2010; Traugott, 2012; Yeager et al., 2019). There should be greater use of representative, probability-based

samples in studying implicit bias. Such samples, constructed within the framework of inferential statistics, are necessary to make valid and reliable inferences to a broader population within a known and calculable margin of sampling error, including design effects. Comparisons of probability and convenience sampling in the literature is focused on discrepancies between measurements of opinions and knowledge of Americans or other population groups in separate administrations of similar questions. It stands to reason that if such issues are observed in this domain, they may also impact basic psychological research as described in this report. To our knowledge, only a handful of studies have attempted to explore implicit measures of attitudes in representative samples of Americans. Moreover, echoing themes in the overview, researchers using implicit measures to study attitudes should acknowledge that their work is being used by policy makers and the general public who assume the basic psychological findings would replicate in samples of American adults and that such findings are important enough to cause alarm about the prevalence and effects of implicit bias, and to inspire organizations to build education and attitude change campaigns in response to them. Representative samples are particularly important for identifying expected effect sizes of postulated outcomes in the general population.

(11) The last two decades of neuroscience research has produced a growing number of studies that suggest that various psychological phenomena are produced by predictive processes in the brain (Hutchinson & Barrett, 2019). Actions, and their accompanying experiences, begin as top-down representations in the brain, fashioned from past experiences that are tested against the state of the world. According to a predictive processing approach, neural predictions (constructed from past experiences) are thought to be a continuously changing filter through which sensory inputs are processed, influencing the relevance of those inputs, effectively deciding which sensory features warrant further processing and action. Once prediction errors are sufficiently minimized, these “inferences” become the brain’s account of what caused the sensations in the first place, effectively categorizing the sensations so that they are meaningful. This approach to understanding processing at the neural level places important constraints on, and suggests ways of understanding the nature of attitudes, assessed with both implicit and explicit measures, as well as their relation to one another. Conducting such work would importantly inform our understanding of these constructs.

(12) Attention to and extension of the developmental literature on implicit attitude measures in children might helpfully inform the broader literature in the field. Research that explores changes as a function of age (with children or with adults) has the potential to help us better understand both the lower-level cognitive and higher-level sociological contributions to performance on these tasks. For example, using age-appropriate versions of the IAT, Baron and Banaji (2006; see also Dunham, Baron, & Banaji, 2006) showed significant bias among children as young as 6 years old. In fact, the magnitude of bias in these children was statistically equivalent to bias in adults. Using a different implicit measure with children between the ages of 9 and 15, Degner and Wentura (2010) showed a very different pattern: at the age of 9, bias was not significant. But bias increased gradually over the course of adolescence (interestingly, during this time children may become more sensitive to social norms, e.g., Hirschfeld, 1996). Degner and Wentura went on to show that among young children, the magnitude of bias on implicit measures depends heavily on the cognitive

demands imposed by the task (this is a critical issue in developmental work, see Crookes & McKone, 2009). On tasks like the IAT, which force the participant to categorize faces by race, young children seem to show pronounced bias. On tasks that do not force categorization, bias may not emerge until later in development. At the other end of the age continuum, work with older participants reinforces the importance of cognitive ability in performance on implicit measures (e.g., Gonsalkorale, Sherman, & Klauer, 2009; Stewart, et al., 2009). As these studies demonstrate, developmental work may help the field disentangle the factors that influence measures of bias and should therefore be encouraged.

(13) Prejudice and discrimination were initially conceived as the product of animosity toward an outgroup (e.g., Allport, Clark, & Pettigrew, 1954). In fact, however, while they can be driven by outgroup negativity, they can also emerge as the product of ingroup positivity, or both (Brewer, 1999). Unfortunately, the most widely used implicit measures of attitudes do not readily allow researchers to separate ingroup positivity from outgroup negativity. For example, the Implicit Association Test is inherently relative, pitting Us/Good and Them/Bad against the opposite set of associations. Nonetheless, there are numerous measures that allow the implicit assessment of bias toward a group in isolation – including variants of the IAT itself (the Single Category IAT; Karpinski & Steinman, 2006). Greater use of such measures would allow assessment of whether implicitly measured ingroup positivity is sufficient, in the absence of implicitly measured outgroup negativity, to generate discriminatory outcomes. Such findings could potentially play an important role in national debates about the underlying causes of prejudice and discrimination and the best ways to combat them. That is, it is necessary to understand the basis of the bias in order to develop methods to change that bias.

## **VIII. A Caution About the Widespread Use of Implicit Bias in Popular Discourse and Policy Making**

Beyond the points made above, a consideration of implicit bias should include an assessment of the resonance and impacts of research claims in society at large. The concept of implicit bias has become firmly entrenched in American society. Government institutions, corporations, and professional associations require or recommend training on implicit bias for their employees and members to recognize and overcome their presumed implicit biases. Information disseminated by researchers, the media, government institutions, public officials, and companies assert that such biases are common, can be measured reliably, influence behavior, and are susceptible to intervention.

Yet, as discussed, these claims are less well-supported by empirical evidence than many people think. Presenters at the NSF Conference on Implicit Bias raised questions regarding the definition and independent predictive power of implicit measures of attitudes, the validity and reliability of these measures (especially in assessing individuals), the consistency and extent of their influence on behavior, and whether and how training can help to overcome implicit bias.

A representative national survey conducted as a follow-up to the NSF conference found that broad majorities of Americans think unconscious biases are prevalent, influence behavior, and can be

mitigated through training, in line with many representations in the public sphere. The public sees unconscious biases as more prevalent than biases that are consciously held and as worthy of mitigation efforts by businesses and government (see Langer, Baskakova, Krosnick, & De Jong, 2019) for more detail. Self-reported exposure to information about implicit bias relates to these views. These attitudes have policy consequences. Despite uncertain empirical evidence of the effectiveness of training to overcome unconscious bias, three-quarters of Americans see such programs as worthwhile. Just as many Americans support spending public funds on unconscious bias training for the police in their communities. And 6 in 10 think such training would in fact change police behavior (albeit without a great deal of confidence in this outcome). These optimistic public beliefs about the value of implicit bias training may ultimately prove to be correct, but unfortunately, at present, the evidence is weak. Thus, more research is needed.

Given the pernicious role of prejudice in society, it is essential for social scientists to devote their best efforts and practices to understanding the causes and effects of bias, both implicit and explicit; to developing, to the extent possible, empirically validated methods of addressing bias; and to communicating their findings accurately and well to policymakers and the public. The current disconnect between verified research claims and public understandings underscores the need for new, richer independent research into the meaning and measurement of implicit bias and its demonstrable impacts and treatability.

## **IX. Conclusion**

Over the past two decades, social psychologists have identified a new and exciting area of inquiry – implicit bias. We have defined this phenomenon above and provided a brief review of what knowledge has emerged on this topic, what is uncertain, and what remains to be done. The phenomenon of implicit bias in some form is real, but the research is still in its infancy. For example, it is unclear what kind of implicit bias (see definitions above) is most pervasive and problematic, how to best assess the presence of such bias, what conditions moderate implicit bias, and how best to address its negative consequences.

## **X. References**

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888.
- Allport, G. (1954). *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.
- Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience, 15*(10), 670.
- Avila, M., Parkin, H., & Galoostian, S. (2019) "\$16.7 Million To Save One Reputation: How Starbucks Responded Amidst a Racial Sensitivity Crisis," Pepperdine Journal of Communication Research: Vol. 7 , Article 4.  
Available at: <https://digitalcommons.pepperdine.edu/pjcr/vol7/iss1/4>

- Axt, J. R., Nguyen, H., & Nosek, B. A. (2018). The Judgment Bias Task: A flexible method for assessing individual differences in social judgment biases. *Journal of Experimental Social Psychology*, 76, 337-355.
- Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, 46, 668–688. doi: 10.3758/s13428-013-0410-6
- 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.
- 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(1), 53-58.
- Baumgartner, F. R., Epp, D. A., & Love, B. (2014). Police searches of Black and White motorists. *UNC-Chapel Hill, Department of Political Science*.
- Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin*, 21(1), 21-31.
- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94(3), 567.
- Bobo, L., Kluegel, J. R., & Smith, R. A. (1997). Laissez-faire racism: The crystallization of a kinder, gentler, antiblack ideology. *Racial attitudes in the 1990s: Continuity and change*, 15, 23-25.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, 55, 429-444.
- Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding implicit bias: Putting the Criticism in Perspective. *Pacific Philosophical Quarterly*, 101, 276-307.
- Cacioppo, J.T. and Berntson, G.G. (1994) Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401–423.
- Calanchini, J., Lai, C. K., & Klauer, K. C. (in press). Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *Journal of Personality and Social Psychology*.
- Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass*, 7, 654-667.

- Callegaro, M., Villar, A., Yeager, D. S., & Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples. In M. Callegaro, R. Baker, J. Bethlehem, A.S. Goritz & J.A. Krosnick (Eds.) *Online Panel Research: A Data Quality Perspective* (pp. 23-53). Hoboken, NJ: John Wiley & Sons.
- Cohn, N., & Quealy, K. (2020, June 10). How public opinion has moved on Black Lives Matter. *New York Times*. Retrieved from: <https://www.nytimes.com/interactive/2020/06/10/upshot/black-lives-matter-attitudes.html>
- Connor, P., & Evers, E.R.K. (2020). The bias of individuals (in crowds): Why implicit bias is probably a noisily measured individual level construct. *Perspectives on Psychological Science*.
- 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(6), 1314.
- Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, 92(6), 1006.
- Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, 111(2), 219-247.
- Darity Jr, W., Hamilton, D., Paul, M., Aja, A., Price, A., Moore, A., & Chiopris, C. (2018). What we get wrong about closing the racial wealth gap. *Samuel DuBois Cook Center on Social Equity and Insight Center for Community Economic Development*.
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135, 347–368.
- Degner, J., & Wentura, D. (2010). Automatic prejudice in childhood and early adolescence. *Journal of Personality and Social Psychology*, 98(3), 356.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Dobbin, F., & Kalev, A. (2018). Why Doesn't Diversity Training Work? The Challenge for Industry and Academia. *Anthropology Now*, 10(2), 48-55.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, 22(1), 22-37.
- Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination

of the relationship between measures of intergroup bias. *Blackwell Handbook of Social Psychology: Intergroup Processes*, 4, 175-197.

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(5), 510-540.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, 77(5), 1268-1281.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23(3), 316-326.

Eberhardt, J. L. (2020). *Biased: Uncovering the hidden prejudice that shapes what we see, think and do*. Penguin Books: New York, NY.

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(6), 876.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?. *Journal of Personality and Social Psychology*, 69(6), 1013.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research. *Journal of Business Venturing* 5, 177-189.

Fishbein, M., & Ajzen, I. (1974). Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59-74.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226-241.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 23-30.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, 45(2), 410-414.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509-517.

- Greenwald, A. G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Greenwald, A.G., & Lai, C.K. (2020). Implicit social cognition. *Annual Review of Psychology*, 71, 419-445.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Task. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- 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(1), 17.
- Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy*, 9(1), 241-253.
- Han, A.H., Czellar, S., Olson, M.A., & Fazio. R. H. (2010). Malleability of attitudes or malleability of the IAT. *Journal of Experimental Social Psychology*, 46, 286-298.
- Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, 9, 393-401.
- Hirschfeld, L. A. (1996). *Race in the Making*. Cambridge: MIT Press.
- Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*. doi: 10.1177/0963721419831992
- Johnson, I., Petty, R. E., Briñol, P., & See. Y.M. (2017). Persuasive message scrutiny as a function of implicit-explicit discrepancies in racial attitudes. *Journal of Experimental Social Psychology*, 70, 222-234.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76(5), 349.
- Jordan, C. H., Whitfield, M., & Zeigler-Hill, V. (2007). Intuition and the correspondence between implicit and explicit self-esteem. *Journal of Personality and Social Psychology*, 93(6), 1067.
- Kalmoe, N. P., & Piston, S. (2013). Is implicit prejudice against blacks politically consequential? Evidence from the AMP. *Public Opinion Quarterly*, 77(1), 305-322.

- Kang, J., Bennett, M., Carbado, D., & Casey, P. (2011). Implicit bias in the courtroom. *UCLA Law Review*, 59, 1124-1186.
- 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.
- Kihlstrom, J. F. (2004). Implicit methods in social psychology. *The Sage Handbook of Methods in Social Psychology*, 195-212.
- Kinder, D. R., and Ryan, T. J. (2017). Prejudice and Politics Re-Examined: The Political Significance of Implicit Racial Bias. *Political Science Research and Methods*, 2, 241-259.
- Kinder, D. R., Sanders, L. M., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*. Chicago: University of Chicago Press.
- Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40(3), 414.
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., & Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*.
- Langer, G., Baskakova, Y., Krosnick, J.A. & De Jong, A. (2019). Public Attitudes on Implicit Bias. *Manuscript submitted for publication*.
- 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 and Medicine*, 170, 220-227.
- Lindgren, K. P., Baldwin, S. A., Olin, C. C., Wiers, R. W., Teachman, B. A., Norris, J., Kaysen, D., & Neighbors, C. (2018). Evaluating within-person change in implicit measures of alcohol associations: Increases in alcohol associations predict increases in drinking risk and vice versa. *Alcohol and Alcoholism*, 53, 386–393.
- Loersch, C., McCaslin, M. J., & Petty, R. E. (2011). Exploring the impact of social judgeability concerns on the interplay of associative and deliberative attitude processes. *Journal of Experimental Social Psychology*, 47, 1029-1032.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81(5), 842.
- MacInnis, B., Krosnick, J.A., Ho, A.S., & Cho, M. (2018) The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly*, 82(4), 707–744.

- Malhotra, N., & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. *Political Analysis*, 15(3), 286-323.
- McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin*, 9(4), 551-558.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64(5), 723.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social cognition*, 19(6), 625-666.
- 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(5), 653.
- Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science & Medicine*, 181, 191–198.
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76(6), 972.
- Pasek, J., & Krosnick, J. A. (2010). Measuring intent to participate and participation in the 2010 census and their correlates and trends: Comparisons of RDD telephone and non-probability sample Internet survey data. Statistical Research Division of the US Census Bureau, Survey Methodology Study Series, 2010(15).
- Pasek, J., Tahk, A., Lelkes, Y., Krosnick, J. A., Payne, B. K., Akhtar, O., & Tompson, T. (2009). Determinants of turnout and candidate choice in the 2008 US presidential election: Illuminating the impact of racial prejudice and other considerations. *Public Opinion Quarterly*, 73(5), 943-994.
- Payne, B.K., Vuletich, H.A. & Lundberg, K.B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28, 233-248.

- 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.
- 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.
- 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*(2), 367-374.
- 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.
- Pérez, E. O. (2010). Explicit evidence on the import of implicit attitudes: The IAT and immigration policy judgments. *Political Behavior, 32*(4), 517-545.
- Petty, R. E., Fazio, R. H., & Briñol, P. (Eds.) (2009). *Attitudes: Insights from the new implicit measures*. New York, NY: Psychology Press.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, Iowa: William C. Brown.
- Petty, R. E., & Krosnick, J. A. (Eds.) (1995). *Attitude strength: Antecedents and consequences*. Mahwah, NJ: Erlbaum Associates.
- Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: an exploration of the PAST model. *Journal of Personality and Social Psychology, 90*(1), 21.
- Petty, R. E., Wheeler, S. C., & Tormala, Z. L. (2003). Persuasion and attitude change. In T. Millon & M. J. Lerner (Eds.), *Handbook of Psychology: Volume 5: Personality and Social Psychology* (pp. 353-382). Hoboken, NJ: John Wiley & Sons.
- Phelps, E. A., O'Connor, K.J., Cunningham, W.A., Funayama, E.E., 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.
- Piston, S. (2010). How explicit racial prejudice hurt Obama in the 2008 election. *Political Behavior, 32*(4), 431-451.
- Proshansky, H. M. (1943). A projective method for the study of attitudes. *Journal of Abnormal and Social Psychology, 38*(3), 393.

- 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.
- Schimmack, U. (2020). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science*.
- Schneider, I. K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., & Koole, S. L. (2015). The path of ambivalence: Tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology*, 6, 996.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, 84(1), 1.
- Sears, D. O. (1998). Racism and politics in the United States. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 76-100). Thousand Oaks, CA, US: Sage Publications, Inc.
- Sears, D. O., & Henry, P. J. (2003). The origins of symbolic racism. *Journal of Personality and Social Psychology*, 85(2), 259.
- Sears, D. O., & Kinder, D. R. (1971). *Racial tension and voting in Los Angeles* (Vol. 156). Institute of Government and Public Affairs, University of California.
- Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, 39, 75-82.
- Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Duinn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, 24(11), 1139-1152.
- 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.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662.
- Tesler, M., & Sears, D. O. (2010). *Obama's race: The 2008 election and the dream of a post-racial America*. Chicago: University of Chicago Press.
- Traugott M. (2012) Methodological Trends and Controversies in the Media's Use of Opinion Polls. In: Holtz-Bacha C., Strömbäck J. (eds) *Opinion Polls and the Media*. Palgrave Macmillan, London
- Vallacher, R. R., Nowak, A., Froehlich, M., & Rockloff, M. (2002). The dynamics of self-

- evaluation. *Personality and Social Psychology Review*, 6(4), 370-379.
- Vallacher, R.R., Nowak, A., & Kaufman, J. (1994). Intrinsic dynamics of social judgment. *Journal of Personality and Social Psychology*, 67, 20-34.
- Vargas, P. T., von Hippel, W., & Petty, R. E. (2004). Using “partially structured” attitude measures to enhance the attitude-behavior relationship. *Personality and Social Psychology Bulletin*, 30, 197-211.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The Linguistic Intergroup Bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, 33, 490-509.
- von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice among the elderly. *Personality and Social Psychology Bulletin*, 26, 523-532.
- Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 25(4), 41–78.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.
- Wittenbrink, B., Judd, C.M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815-827.
- Wittenbrink, B., & Schwarz, N. (Eds.). (2007). *Implicit measures of attitudes*. Guilford Press.
- Yeager, D.S., Krosnick, J.A., Visser, P.S., Holbrook, A.L., & Tahk, A.M. (2019). Moderation of classic social psychological effects by demographics in the U.S. adult population: New opportunities for theoretical advancement. *Journal of Personality and Social Psychology*, 117(6), e84-e99.
- Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, 90(3), 553.