10

Popularity as a Poor Proxy for Utility
The Case of Implicit Prejudice

Gregory Mitchell and Philip E. Tetlock

Introduction

It is difficult to find a psychological construct that has moved faster from psychology journals into other academic disciplines, newspaper editorials, courtrooms, boardrooms, and popular consciousness than has the implicit prejudice construct. The first reference to the term "implicit prejudice" in the PsycINFO database appears in a source less than 20 years old (Wittenbrink, Judd, & Park, 1997). A Google search for "implicit prejudice" between the years 1800 and 1990 returns only six hits, while the same search for the years 1991–2016 returns over 8400 hits. Google Scholar returns 46 hits for the phrase "implicit prejudice" in sources published between 1800 and 1990, but over 3700 hits for sources published after 1990. The 1998 article introducing the implicit association test (IAT) (Greenwald, McGhee, & Schwartz, 1998), which is now the most popular method for studying implicit prejudice, has already been cited more than 3700 times in PsycINFO, 3400 times in the Web of Science database, and 7500 times in Google Scholar.

It is also difficult to find a psychological construct that is so popular yet so misunderstood and lacking in theoretical and practical payoff. Scholarly discussions of prejudice fail to agree on how implicit prejudice connects to other forms of prejudice; it is unclear whether different measures of implicit prejudice measure the same thing; the meaning of "implicit" in the phrase "implicit prejudice" is contested; and implicit measures of prejudice are no better at predicting behavior, even "microaggression" (small, barely visible slights), than are traditional explicit measures of prejudice.

How can the grand popularity of the implicit prejudice construct be reconciled with the meager theoretical and practical accomplishments of the research

Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions, First Edition.
Edited by Scott O. Lilienfeld and Irwin D. Waldman.

program? Although the implicit prejudice construct is perhaps unique in how fast it gained so much attention, we posit that it is not unique among social science ideas in how it gained its popularity. The attention paid to the implicit prejudice construct illustrates how success in social science can depend less on theoretical clarity or predictive success and more on how skillfully like-minded researchers can use a paradigm to generate statistically significant but substantively insignificant results that they can then package into sound bites that support a particular worldview or political agenda. Concordance with pet theories or political sympathies is not, however, sufficient: many research findings from psychology support a liberal agenda and many economic theories support a conservative agenda (for evidence on the political imbalances in these fields, see Gross, 2013), but few find real fame or wield much influence outside their narrow academic domains. To find real fame, the social scientists behind the construct must also find allies among scholars outside of their own discipline, funding agencies, members of the press, and elites who can sway corporate boards, judges, legislators, and bureaucrats.

Implicit prejudice researchers, particularly the creators of the IAT, have been remarkably adept at forging alliances and popularizing the notion of implicit prejudice (see Chapter 9). For many scholars and public intellectuals (for a recent example, see http://www.nytimes.com/2014/08/28/opinion/nicholas-kristof-is-everyone-a-little-bit-racist.html?_r=0), the implicit prejudice construct has become the go-to explanation for all manner of ills suffered by one group at the hands of another, even when the groups consist of monkeys rather than human beings (see Kang, 2012, relying on the now-retracted Mahajan et al., 2011 for the claim that implicit bias is hard-wired into primate brains; see Mahajan et al., 2014, for the retraction). When the actor who played Kramer on Seinfeld hurls a racial epithet at a heckler during a comedy show (Shermer, 2006); when the cover of the *New Yorker* portrays Barack Obama as a militant Muslim (Banaji, 2008); when Barack Obama beats Hillary Clinton in the 2008 Democratic presidential primary (Kristof, 2008); when surveys find a majority of whites opining that blacks overestimate the frequency of discrimination (Blow, 2009); when a black teen is shot and killed by a neighborhood watch zealot (Feingold & Lorang, 2012; Reeves, 2012); when we try to understand why philosophy departments have so few female professors (Crouch & Schwartzman, 2012); indeed, when virtually any racially or sexually charged event occurs or any disparity in group outcomes materializes, we can depend on the usual-suspect public intellectuals to discern the workings of implicit prejudice.

Of course, there is a body of psychological research behind this implicit prejudice meme, but in this chapter we explain why that body is inadequate to support the uses to which it is being put. Before doing so, however, we discuss how this meme was manufactured, detailing how key psychologists marketed the core ideas. We also discuss several reasons why the implicit prejudice construct is in need of renovation, why the implicit prejudice meme should be retired, and why it is so difficult to combat politically seductive ideas within social psychology (see Chapter 9).

Creating the Implicit Prejudice Meme

The history of the implicit prejudice construct can be divided into two eras: (1) *the pre-IAT era*, in which psychologists developed indirect measures of prejudice aimed at overcoming response biases and began examining automatic processes that may contribute to contemporary forms of prejudice; and (2) *the post-IAT era*, in which implicit prejudice became synonymous in public discussions (and in many academic ones) with widespread unconscious prejudices that are harder to spot than old-fashioned explicit prejudices and that supposedly regularly infect intergroup interactions. A 1998 story in *Psychology Today* on the IAT heralds this new era: "Psychologists once believed that only bigoted people used stereotypes. Now the study of unconscious bias is revealing the unsettling truth: We all use stereotypes, all the time, without knowing it. We have met the enemy of equality, and the enemy is us" (Paul, 1998). The *Psychology Today* story analogizes the IAT to the microscope: just as the microscope allowed biologists to see the previously undetectable viruses that lead to bodily ills, the IAT allows psychologists to see the previously undetectable mental forces that lead to social ills. In a presentation at the 2001 convention of the American Psychological Society (now the Association for Psychological Science), Dr. Banaji embraced a similar view and described the IAT as ushering in a third great scientific revolution to follow the Copernican and Darwinian revolutions (Kester, 2001). The 1998 *Psychology Today* article also gave voice to the now-common idea that it is more difficult to avoid the negative effects of implicit as opposed to explicit prejudice: "[Our] internal censor successfully restrains overtly biased responses. But there's still the danger of leakage, which often shows up in non-verbal behavior: our expressions, our stance, how far away we stand, how much eye contact we make" (Paul, 1998).

Before the IAT arrived on the scene, the ideas of automatic stereotyping and unintentional prejudice were often discussed among psychologists, and sometimes outside of psychology. But, the idea that prejudice operates pervasively and routinely at subconscious levels; *and* that this implicit prejudice contaminates a wide array of judgments, decisions, and behaviors; *and* that this pernicious hidden bias can be reliably measured – these ideas took root with the marketing of the IAT, which supposedly documents widespread implicit preferences for majority groups over minority groups (even among members of the minority groups) that are more predictive of behavior than explicitly measured prejudice.

A review of the public record leaves little doubt that the seminal event in the public history of the implicit prejudice construct was the introduction of the IAT in 1998, followed closely by the launching of the Project Implicit website in that same year (Banaji & Greenwald, 2013). Before 1998, few public discussions of implicit prejudice are found: Google returns 18 results for the the phrase "implicit prejudice" between 1800 and 1997. After 1998, references to implicit prejudice skyrocket: Google returns over 8000 sources using the phrase "implicit prejudice" since the beginning of 1998. As of 2013, over 14 million IATs had been taken through the Project Implicit website, and the website receives over 20,000 new visitors per week

(Banaji & Greenwald, 2013). Many visitors are directed there by other websites that link to the Project Implicit website (e.g., a WooRank search finds over 5000 websites referring visitors to the Project Implicit site). Variations on the American Project Implicit website have been launched in 39 countries in 24 different languages (Banaji & Greenwald, 2013). IATs taken through the Project Implicit websites serve as a key source of data for many of the published IAT studies.

A review of the history of dissemination of information to the public about IAT research yields three striking findings. First, public dissemination of information about the IAT and its significance began shortly after the IAT's official birth. Drs. Anthony Greenwald and Mahzarin Banaji, creators of the IAT along with Dr. Brian Nosek, held a press conference in 1998 to publicize the IAT and announce the launching of the Project Implicit website. At the press conference, the race IAT was said to reveal unconscious prejudice that affects "90–95 percent of people," but Greenwald and Banaji expressed hope that "... the test ultimately can have a positive effect despite its initial negative impact. The same test that reveals these roots of prejudice has the potential to let people learn more about and perhaps overcome these disturbing inclinations" (http://www.washington.edu/news/1998/09/29/roots-of-unconscious-prejudice-affect-90-to-95-percent-of-people-psychologists-demonstrate-at-press-conference). Publicity surrounding the publication of the first IAT study led to stories in *Psychology Today* (Paul, 1998), the Associated Press (Tibbits, 1998), and the *New York Times* (Goode, 1998), with these articles generating further articles.[1] According to Factiva, in 1998, at least 19 stories on the IAT were published in major newspapers and wire services, with many college and local newspapers in turn picking up and reporting these stories.

The second striking finding is the breadth of the marketing effort, which has been sustained over several years and has been multifaceted, involving multiple media outlets and multiple disciplines. The 1998 print stories were the first of many: Factiva's newspaper and newswire databases presently contain over 400 stories containing the phrase "implicit association test" and over 1200 stories containing the phrase "implicit bias." Magazines have published a number of stories on the IAT as well (e.g., *Newsweek* has published at least three stories discussing the IAT since 2008). Many of these stories encourage readers to visit the Project Implicit website, which provides additional educational information on implicit bias.

Television, radio, and Internet media have paid considerable attention to the IAT as well, often with the help of the IAT's creators. In November 1998, Greenwald appeared on an NBC News segment, demonstrating the IAT as a measure of unconscious prejudice, and in March 2000, NBC's *Dateline* program aired a segment on the IAT, and again in 2007, following derogatory comments by Don Imus about female basketball players. In the *Dateline* episode, Banaji stated that the IAT reveals how "fair are we being when we judge a person," and Greenwald gave an example of the wrongful shooting of a black suspect by police as an example of how the bias measured by the IAT can affect behavior. In 2002, Greenwald appeared again on an NBC *Nightly News* segment, relating implicit bias as measured by the IAT to wrongful police shootings. In 2006, Greenwald appeared on a segment of ABC's *20/20*

news show discussing the IAT, and later that same year Banaji appeared on Paula Zahn's CNN show discussing the IAT. In March 2013, Greenwald appeared on PBS's *Tavis Smiley Show* to discuss IAT research. The IAT has even made an appearance on *Fox News* when, in 2005, a guest on Bill O'Reilly's show, while discussing the execution of Tookie Williams by the state of California, referred to the IAT as a test that "demonstrates that we infuse bias into our decision-making processes when we evaluate evidence."

National Public Radio ("NPR") has aired several stories on the IAT. For instance, NPR used IAT research in its coverage of the incident at a comedy club involving racist remarks by Michael Richards ("Kramer" from *Seinfeld*) in 2006, indicating that implicit bias may have played a role in the incident. Also, in an NPR story on the role of race in the 2012 presidential election, Greenwald described the implicit bias construct for the audience and discussed the behavioral effects of implicit bias: "But people aren't actually aware that they have this. They often explicitly reject it. They certainly don't want to have it. But nevertheless, it can act on them, and it can affect their behavior. It can produce discomfort in interracial interactions, and that's something that all by itself is likely to produce some unintended discrimination." In 2008, Brian Nosek appeared on NPR's *Talk of the Nation* program in a segment devoted to examining "tests that can reveal your hidden bigotry." Nosek explained that, while these hidden biases may not lead to extreme examples of racism, such as KKK-type assaults, they are likely to lead to subtle behaviors that can have adverse effects, such as causing discomfort in employment interviews.

News and opinion websites, as well as many blogs and educational websites, have also given extensive coverage to IAT research. For instance, a search of the *Huffington Post* site for "implicit association test" yields over 70 hits, and the Southern Poverty Law Center's tolerance.org website has a page to "Test Yourself for Hidden Bias" that discusses implicit prejudice and directs readers to the Project Implicit website.

The implicit prejudice meme has also been advanced by popular science writers, most notably Malcolm Gladwell in his 2005 book *Blink*. In addition to devoting a section of *Blink* to implicit bias and discrimination (where he gives a hypothetical example of a white interviewer whose implicit prejudice leads to subtle discrimination against a black interviewee), shortly after publication of *Blink*, Gladwell appeared on Anderson Cooper's CNN show and linked implicit bias to the shooting of Amadou Diallo and to price discrimination against black car buyers, further solidifying the implicit-prejudice-leads-to-discrimination meme.[2] Popular science writer Shankar Vedantam also published a book devoted to discussing what he called "unconscious prejudices – subtle cognitive errors that lay beneath the realm of awareness" (Vedantam, 2010, p. 3). IAT research figures prominently in Vedantam's book, and he invokes unconscious racism and unconscious sexism to explain a wide variety of events – from George Allen's now infamous "macacca" comment during the 2008 senate race in Virginia (when Allen referred to an Indian-descent volunteer of the opposing campaign as "macaca," a term sometimes used to refer to a monkey), to Hillary Clinton's showing in the 2008 presidential primary, to racial disparities in the death penalty, and male–female pay differentials.

Banaji and Greenwald recently added their own book popularizing IAT research (Banaji & Greenwald, 2013). In *Blindspot: Hidden Biases of Good People*, the reader is assured that an objective account of the research is coming: "we have chosen to stick closely to the evidence, especially experiments whose conclusions reflect widely shared consensus among experts" (Banaji & Greenwald, 2013, p. xv). The implicit-prejudice-leads-to-discrimination meme is presented as part of this fact-based consensus:

> the automatic White preference expressed on the Race IAT is now established as signaling discriminatory behavior. It predicts discriminatory behavior even among research participants who earnestly (and, we believe, honestly) espouse egalitarian beliefs. That last statement may sound like a self-contradiction, but it's an empirical truth. Among research participants who describe themselves as racially egalitarian, the Race IAT has been shown, reliably and repeatedly, to predict discriminatory behavior that was observed in the research (Banaji & Greenwald, 2013, p. 47).

Later, in an appendix to the book, Banaji and Greenwald discuss inequalities in housing, hiring, health care, and criminal justice outcomes, and then write that "it is reasonable to conclude not only that implicit bias is a cause of Black disadvantage but also that it plausibly plays a greater role than does explicit bias in explaining the discrimination that contributes to Black disadvantage" (Banaji & Greenwald, 2013, p. 209).

In addition to seeking to influence public views on the meaning and prevalence of prejudice through *Blindspot*, through presentations to the general public and academic audiences, and through interactions with the media, Greenwald and his colleagues have sought to influence how courts and juries think about prejudice and discrimination. "The central idea is to use the energy generated by research on unconscious forms of prejudice to understand and challenge the notion of intentionality in the law," Banaji told a reporter with the *Harvard Gazette* (Potier, 2004). In describing to the reporter why this project to change the law was so important, Greenwald used the Amadou Diallo case as an example of the behavioral consequences of implicit bias (Potier, 2004). Greenwald has now appeared as an expert witness in several legal cases (in some of these cases, the authors of this chapter have offered responsive reports discussing the limits of the IAT research), and Banaji testified about the possible influence of implicit bias on jurors in a death penalty case in New Hampshire. Greenwald has given presentations at American Bar Association conferences aimed at educating lawyers on possible legal implications of the IAT research, and Greenwald and Banaji have both co-authored papers with legal scholars for legal audiences (e.g., Greenwald & Krieger, 2006; Kang & Banaji, 2006; Kang et al., 2012). One of their legal collaborators, Professor Jerry Kang, frequently gives talks to law firms and companies about the dangers of implicit bias (see http://jerrykang.net/talk/implicit-bias-talks), and Kang developed a primer on implicit bias for use by the National Center for State Courts as part of a program to educate state court judges and other personnel on the dangers of implicit bias

(see http://www.ncsc.org/~/media/files/pdf/topics/gender%20and%20racial%20fairness/kangibprimer.ashx). Kang also gave a TEDx talk that should help further spread the implicit prejudice meme (see the video at https://www.youtube.com/watch?v=9VGbwNI6Ssk&feature=youtu.be).

The IAT's creators have also marketed IAT research to Fortune 500 companies. Many Fortune 500 companies now include discussions of the IAT and implicit bias in their diversity training (Lublin, 2014), with a good bit of these consultations being provided by Project Implicit, Inc., a non-profit organization started by Greenwald, Banaji, and Nosek to provide paid consulting services to organizations (among other services).[3] As shown on publicly available tax returns, Project Implicit, Inc. has earned several hundred thousand dollars from its consulting services, with substantial portions of this money being given as grants to IAT researchers.

Funding from Project Implicit, Inc. is only part of the substantial resources that have been provided to develop and promote IAT research. Federal grant agencies were strong supporters of the IAT research program from its beginning, with Greenwald, Banaji, and Nosek all having received federal grants to perform research into implicit social cognition (Greenwald received a grant as early as 1992 to perform research on implicit prejudice). This funding not only enabled much data collection but also the training of many graduate students and postdoctoral fellows who now use the IAT to study implicit prejudice and other topics. Graduates of the labs of Greenwald and Banaji are now ardent defenders of IAT research and of the view that implicit prejudice is a force that must be reckoned with if society is to address its many inequalities (see, e.g., Blasi & Jost, 2006; Jost et al., 2009).

The third striking fact evident from a review of the public history of IAT research is the boldness of the claims that have been made about the meaning and implications of IAT research, even before a single published study had linked scores on an IAT to any behaviors. Indeed, one can view the implicit prejudice meme as a direct descendant of early and continuing proclamations by IAT researchers and their affiliates about the behavioral potency of bias as measured by the IAT. As noted in the preceding text, as early as 2000, Greenwald linked implicit bias as measured on the IAT with acts of wrongful police shooting and workplace discrimination. One consistent theme in public discussions of IAT research, as demonstrated in Nosek's comments on the *Talk of the Nation* show and Malcolm Gladwell's comments in *Blink* and on CNN, has been that implicit prejudice leads to snap judgments and uncomfortable interpersonal interactions that adversely affect women and minorities, in encounters with police, in employment interviews, in workplace teams, and in other situations (see, e.g., Chugh, 2004, for a discussion of the subtle biasing effects implicit prejudice might have in work settings). But we see the implicit prejudice meme broadening to encompass deliberative judgments and decisions and macro-level behaviors, as in the appendix to *Blindspot*. Currently, on the frequently asked questions page of the Project Implicit website, visitors are presented with the question "If my IAT shows that I have an implicit preference for one group over another, does that mean I am prejudiced?" and are informed that "[t]he IAT shows biases that are not endorsed and that may even be contradictory to what one

consciously believes. So, no, we would not say that such people are prejudiced. *It is important to know, however, that implicit biases can predict behavior. When we relax our active efforts to be egalitarian, our implicit biases can lead to discriminatory behavior, so it is critical to be mindful of this possibility if we want to avoid prejudice and discrimination*" (https://implicit.harvard.edu/implicit/faqs.html#faq3; emphasis added). The unmistakable message from Project Implicit and numerous other sources of information is that implicit biases pervade our interpersonal interactions at many levels, and even the best intentions will often not guard against their impacts on behavior.

Intermediaries of social science research have passed this message on to their respective audiences. Captain Gove, of the Hartford Police Department, Connecticut, after seeing a presentation by Jerry Kang, writes in *The Police Chief Magazine* (Gove, 2011) that implicit biases are pervasive and lead to discrimination: "From simple acts of friendliness and inclusion to more consequential acts such as the evaluation of work quality, those who are higher in implicit bias have been shown to display greater discrimination." The CEO of Workforce Answers, a firm that provides legal compliance training to companies, writes that "[e]xperts believe that secret biases – biases that people don't even know they hold – still affect their personal and professional decisions. This 'implicit bias' is thought to be a reason for much discrimination" (Lieber, 2009, p. 93). Law professors writing about discrimination now regularly pay heed to implicit bias and its behavioral effects (e.g., Bagenstos, 2007; Benforado & Hanson, 2008; Garda, 2011; Gomez, 2013; Green, 2010; Levinson & Smith, 2012; Richardson, 2011; Robinson, 2008); public defenders worry that implicit bias adversely affects their clients in many ways (e.g., http://davisvanguard.org/the-role-of-implicit-bias-and-how-it-impacts-cases-like-trayvon-martin/); the National Center for State Courts warns court personnel that implicit bias may affect a judge's sentencing decisions, an employer's hiring decisions, or a police officer's decisions to shoot (http://www.ncsc.org/~/media/Files/PDF/Topics/Gender%20and%20Racial%20Fairness/Implicit%20Bias%20FAQs%20rev.ashx); human resource advisors warn about implicit bias effects on personnel decisions (e.g., Babcock, 2006); universities provide primers to faculty search committees on the dangers of implicit bias (e.g., http://facultyhiring.uoregon.edu/files/2011/05/Best-Man-For-The-Job-How-Bias-Affects-Hiring-qymz6i.pdf); and medical researchers warn doctors about how their implicit biases are contributing to racial disparities in health (e.g., Chapman, Kaatz, & Carnes, 2013). These examples of applications of the IAT research, and the implicit prejudice meme that it supports, are only a handful of the many examples that could be offered.

The dedicated efforts of the IAT researchers, with the assistance of many others, to publicize IAT research and to promote the view that implicit prejudices are an important source of discrimination that must be addressed have been remarkably successful. The implicit prejudice meme appears now to be self-sustaining: it is now so widespread and commonly invoked that new invocations of the meme need merely cite the many prior invocations of the meme, with little attention ever given to the origins of the meme and to whether those origins can actually support the

claims being made. Shankar Vedantam, in his review of *Blindspot* for NPR, concludes that "[Banaji and Greenwald] have revolutionized the scientific study of prejudice in recent decades, and their Implicit Association Test – which measures the speed of people's hidden associations – has been applied to the practice of medicine, law and other fields. Few would doubt its impact, including critics" (http://www.wbur.org/npr/177455764/What-Does-Modern-Prejudice-Look-Like). We cannot speak for others, but the present critics do not doubt the impact of the IAT on public beliefs about implicit prejudice. We do, however, doubt the IAT's theoretical and practical contributions and the value of the implicit prejudice construct.

Deconstructing the Implicit Prejudice Meme

It is our contention that, when the public rhetoric about IAT research is compared to the details of the underlying research, the social and scientific significance of this research becomes much less apparent. To validate this contention, we discuss problems in the formulation and measurement of the implicit prejudice construct itself, and then we move to questions bearing on real-world applications of the construct. On issue after issue, there is little evidence of positive impacts from IAT research: theories and understandings of prejudice have not converged as a result of the IAT research; bold claims about the superior predictive validity of the IAT over explicit measures have been falsified; IAT scores have been found to add practically no explanatory power in studies of discriminatory behavior; and IAT research has not led to new practical solutions to discrimination. Only two indisputable professional contributions have been made by development of the IAT, both of uncertain scientific and social value: (a) the documentation of replicable statistically significant differences in response patterns to opposing attitude objects on the IAT and (b) the facilitation of the publication of journal articles that report these response patterns. The idea that the IAT has opened our eyes to a new form of prejudice that pervades and degrades intergroup interactions should be retired, and the implicit prejudice construct should be subjected to greater theoretical and empirical scrutiny.

Our contention is threatening to those who have made public claims about the scientific and social significance of the IAT research and who benefit professionally and financially from the popularity of IAT research. Disagreements over the scientific merits of the IAT to the side, there is one thing on which proponents and skeptics of the test can agree: many professors have advanced their careers thanks to the IAT (whether serving as advocates or critics of the test), it has spawned a cottage industry of diversity consultants offering unproven implicit bias training programs, and it has given lawyers much to fight about (and bill for) in many lawsuits. As a result, some will be (consciously or unconsciously) motivated to mischaracterize our arguments, question our motives, and cherry-pick favorable results to try to dismiss the evidence we cite, as has already occurred with our past criticisms of the public interpretations and applications of IAT research. For example, the views of Arkes and Tetlock (2004) were likened to the views of the Supreme Court justices

who decided the infamous case of *Plessy v. Ferguson* (Banaji, Nosek, & Greenwald, 2004), and the Mitchell and Tetlock (2006) article casting doubt on legal implications of the IAT research has been described as "predictable political backlash, regrettably laced with ad hominem and strawperson excess" (Lane, Kang, & Banaji, 2007, p. 442). And the fact that we have provided expert consulting services to companies confronted with an expert report prepared for the plaintiffs by an IAT researcher has been offered as evidence of our bias, whereas the expert services provided by the psychologists to whom we respond seem never to provoke contamination concerns about their work.

It would be folly, in an article on implicit bias, to try to convince the reader that self-reported noble, scientific intentions motivate our criticism of the implicit prejudice meme. All we can ask is that the reader try to consider our arguments and evidence with an open mind. In considering our points, keep in mind that we are often repeating or summarizing points made by other researchers who have raised questions about the construct and external validity of the implicit prejudice research. Despite efforts to portray those who raise questions about the IAT as a small group of discontents,[4] the fact is that many researchers have serious questions about the meaning and implications of IAT scores and about the larger implicit prejudice construct.

A few final prefatory comments aimed at preventing mischaracterization and misunderstanding: We do not deny that research into implicit social cognition, and particularly the role of automatic processes in stereotyping and prejudice, has produced some important theoretical insights, and we certainly do not deny that prejudice continues to be an important social problem that contributes to inequalities. We recognize that implicit measures other than the IAT exist and have produced influential findings, but we believe it is indisputable that IAT research serves as the backbone of the implicit prejudice meme. And we believe it is indisputable that existing empirical research, whether based on the IAT or any other implicit measure of prejudice, cannot support the weight of the implicit prejudice meme.

What Is Implicit Prejudice, and Why Don't Its Measures Agree?

Two related themes are repeatedly found in works discussing and seeking to test for the presence of implicit prejudice. First, social psychologists express great skepticism about the accuracy of survey-based estimates of the declining prevalence of prejudicial attitudes and stereotypes due to social desirability pressures on survey respondents. From this perspective, reaction-time-based measures of prejudice, such as the lexical decision task (Wittenbrink et al., 1997) and the IAT (Greenwald et al., 1998), represent an evolution of unobtrusive measures of prejudice that seek to assess prejudice indirectly to avoid the influence of normative pressures (e.g., Crosby, Bromley, & Saxe, 1980; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Second, social psychologists, particularly since the 1990s, have shown renewed faith

in their ability to tap into subconscious influences on judgments, decisions, and behavior. The implicit prejudice construct thus reflects an evolution of views about the nature of attitudes (e.g., Banaji & Greenwald, 1995; Wilson, Lindsey, & Schooler, 2000) and about the influence of automatic psychological processes and their influence on behavior (e.g., Bargh, Chen, & Burrows, 1996).

The interrelated nature of these themes has given rise to one of the fundamental confusions that surrounds the implicit prejudice construct: are the processes encapsulated by the construct implicit (i.e., operating beyond self-awareness and/or conscious control), or is the means of measuring prejudice implicit (i.e., the object of inquiry is unknown to the subjects)? Many works fail to distinguish between these two senses of the modifier "implicit," often using the modifier in both senses (De Houwer & Moors, 2007; Fazio & Olson, 2003). Reflecting this confusion, even experts on prejudice disagree about how to define implicit prejudice and how to describe the underlying psychological processes. For instance, different definitions are offered across chapters in the most recent iteration of the *Handbook of Social Psychology*: in the chapter on "Intergoup Bias," Dovidio and Gaertner (2010, p. 1084) embraced implicitness in reference to the kinds of processes measured, referring to bias as "explicit (overt and intentional) or implicit (involving the spontaneous, frequently automatic, activation of evaluations or beliefs...)," while, in the chapter on "Intergroup Relations," Yzerbyt and Demoulin (2010, pp. 1044–1045) embraced implicitness as referring to the mode of measurement, describing implicit measures as allowing "researchers to assess individuals' levels of prejudice in a way that bypasses their attempts to exert control over their responses and are, therefore, quite distinct from their overt response."

Among those who treat implicit prejudice as primarily about the nature of the measured processes, one finds disagreement about the nature of those processes. Dasgupta and Stout (2012), for instance, wrote that implicit biases sometimes operate beneath awareness and, at other times, individuals are aware of these biases but unable to control them.[5] Contrast Dasgupta and Stout's inclusion of both conscious-but-uncontrollable and unconscious bias under the implicit bias banner with Duckitt's (2003, p. 569) crisp distinction between explicit prejudice as operating at a conscious level and implicit prejudice as operating "in an unconscious and automatic fashion." Hardin and Banaji (2012, p. 16) hedge their bets on the automaticity of implicit prejudice by writing that it "operates ubiquitously in the course of normal workaday information processing, *often* outside of individual awareness, in the absence of personal animus, and *generally* despite individual equanimity and deliberate attempts to avoid prejudice" (emphasis added). Brown (2010) hedged on the nature of both implicit *and* explicit prejudice, defining *explicit prejudice* as "[a] direct form of prejudice, which is usually under the person's control" (p. 283), and *implicit prejudice* as "[a]n indirect form of prejudice which typically is not (much) under the person's control" (p. 285).

Even greater hedging may be in order, however, for research casts doubt on the assumption that respondents to implicit measures fail to appreciate what is being measured and have no control over the measured processes. Bar-Anan and Nosek

(2012) challenged the validity of the Affective Misattribution Procedure ("AMP") as a measure of unconscious processes that might result in intergroup bias (for a response disputing this contention, see Payne et al., 2013), and Hahn and colleagues (Hahn, Judd, Hirsh, & Blair, 2014) presented evidence leading to the same negative conclusion about the IAT as a means of accessing unconscious and inaccessible attitudes (see also Gawronski, Hofmann, & Wilbur, 2006; Gawronski, LeBel, & Peters, 2007).

Lane et al. (2007, p. 429) told their readers that implicit measures of prejudice such as the IAT "bypass the mind's access to conscious cognition" and "tell us something different from self-reported survey-type responses." Yet, it is not even clear that the two most reliable implicit measures of prejudice (the AMP and IAT) really *are* implicit measures, at least not for all respondents (for a broad critique of the role of untested and often unstated assumptions in conjunction with many implicit measures, see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009).

Moreover, it is not clear that implicit and explicit measures tap into different psychological sources, although it is common to portray the measures as if they do (as in Quillian, 2006, and as in the preceding quotation from Lane et al., 2007). As Nosek (2005) discussed, there are two distinct views within the psychological literature on the explicit–implicit relation: (a) the view that explicit and implicit measures assess distinct constructs, and (b) the view that both measure a single attitude construct, with divergence in responses being due to different levels of conscious or controlled processing (see also Fazio, 2007; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Although evidence in support of both views has been offered, according to Greenwald and colleagues (Greenwald, Poehlman, Uhlmann, & Banaji, 2009, p. 32), "the question of single versus dual representations appears empirically irresolvable" (see also Greenwald & Nosek, 2008). If this epistemological stance is correct, then it casts doubt on the many distinctions drawn within and outside academic psychology between explicit and implicit prejudice.

Adding to the confusion about what exactly comprises implicit prejudice, authors sometimes treat implicit prejudice as encompassing many of the different "modern" forms of prejudice that have been posited to contrast with "traditional" forms of prejudice. Hardin and Banaji (2012) dated the "discovery of implicit prejudice" to Devine's (1989) classic paper examining the effects of priming on stereotype activation, and Hardin and Banaji treated all manner of studies aimed at detecting automatic processes of prejudice and stereotyping as falling under the implicit prejudice banner, from aversive racism research to subliminal priming studies to startle response studies to IAT studies. Dasgupta and Stout (2012, p. 400), citing research from a variety of paradigms including IAT and aversive racism research, wrote that "even people who report egalitarian attitudes toward disadvantaged groups may subtly (or implicitly) favor some social groups and be biased against others in ways that are consistent with social stereotypes."

The inclusion of aversive racism and IAT research under the implicit prejudice banner might suggest a commonality of processing, inputs, and effects, but in fact the association-strength theory behind the IAT differs from the conflict theory

behind aversive racism, and the two forms of prejudice are posited to operate differently. Whereas aversive racism is theorized to result from a conflict between automatic cognitive processing and value- and norm-driven conscious opposition to prejudice and discrimination, with bias manifesting itself in pro-in-group behavior under circumstances where we can attribute the behavior to nondiscriminatory factors (Hodson, Dovidio, & Gaertner, 2004), the bias measured by IATs supposedly reflects the strength of associations between an attitude object and attributes, and there is no clear theory about when these associations will and will not be expressed on the IAT or in behavior.[6] Originally, IAT researchers claimed that bias as measured by the IAT would be more likely expressed in "micro-level" and spontaneous behaviors, but recently they have revised that claim (Greenwald et al., 2009). To be sure, IAT researchers discuss moderators of the IAT effect and of the bias–behavior relation, but these moderator relations are empirically rather than theoretically derived. Aversive racism researchers contend that aversive racism is more predictive of in-group favoritism than affirmative out-group mistreatment (Hodson et al., 2004), but the meta-analysis of IAT behavior studies conducted by Greenwald et al. (2009) did not even examine whether in-group favoritism occurred for many of the criterion variables studied (see the supplement to Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013, for detailed discussion of this issue). The follow-up IAT meta-analysis by Oswald et al. (2013) did examine such behaviors, and it found that the race and ethnicity IATs were very poor predictors of expressions of in-group favoritism. These differences in theory and results illustrate the importance of treating these lines of research separately for scientific and applied purposes: the motivating theories and results from each line of research are not interchangeable, and placing both constructs under the broad implicit prejudice label conceals important differences between the research programs.

The need to differentiate among implicit prejudice research paradigms is further illustrated by the fact that different implicit measures of prejudice produce different patterns of results (Gawronski, 2009). Even measures based on similar methods, such as response latency measures aimed at measuring the strength of associations between groups and positive/negative evaluations (e.g., affective priming and the IAT), produce divergent results (Duckitt, 2003; Fazio & Olson, 2003). Correlations among measures are often low, and the measures typically produce different aggregate levels of bias, with IATs typically showing the highest levels. Thus, if a sequential priming procedure leads to an estimate that 50% of white respondents are implicitly prejudiced against blacks, and the race IAT leads to an estimate that 75% of white respondents are implicitly racist, should only the higher estimate be provided to the public? Although the IAT possesses greater test–retest reliability than most other current implicit measures (albeit still at levels well below that desired for applied use purposes), there is no basis for treating bias as measured by the IAT as more "real" or consequential than bias as measured by the Affect Misattribution Procedure or some other sequential priming method, given that the IAT does not correlate more highly or more reliably with judgments, decisions, or behaviors than sequential priming measures (compare the bias–behavior

correlation estimates in Cameron, Brown-Iannuzzi, & Payne, 2012, and Forscher et al., 2016, with those in Greenwald et al., 2009, and Oswald et al., 2013). In light of the mixed results with respect to correlations among implicit measures and between implicit and explicit measures of prejudice, Dovidio and Gaertner (2010, p. 1108) concluded that the "modest relationships among the various measures of bias suggest the need to refine different conceptions of the elements of bias and further delineate the factors that might moderate the relations among these variables."

To make matters worse, debate continues over the degree to which the IAT effect is the product of artifacts as opposed to the strength of associations with attitude object categories. Although public proclamations about the IAT often describe the IAT as measuring implicit or automatic "preferences" for different groups, strongly implying that the associations reflect both personal preferences and have adverse implications for choices implicating these groups, in fact the degree to which the IAT measures negativity toward, versus empathy for, different groups remains disputed, as does the degree to which the IAT measures attitudinal associations versus other salient associations or constructs (see, e.g., Andreychik & Gill, 2012; De Houwer et al., 2009; Han, Czellar, Olson, & Fazio, 2010; Siegel, Dougherty, & Huber, 2012; Siegel, Sigall, & Huber, 2012). The findings of Andreychik and Gill (2012) should be particularly troubling for promoters of the implicit prejudice meme, for if (or when) the IAT measures empathy instead of negative group attitudes, the behavioral implications of IAT scores should be interpreted quite differently: "our results suggest that measures of implicit evaluation, because they fail to detect the difference between empathy-based and prejudice-based associations, do not provide high-fidelity information about attitudes" (Andreychik & Gill, 2012, p. 1092).

Were one to accept as truth the public proclamations made about the revolutionary nature of IAT research, one might believe that there is now widespread agreement about what exactly the IAT measures and about the "implicitness" of the IAT and other measures of implicit prejudice. And one might believe that IAT research has led to a clear, consensual understanding of the nature of implicit prejudice and its relation to explicit prejudice. Those beliefs would be mistaken.

Predictive validity is essential but lacking

Though much debate remains about the nature and proper definition of attitudes, self-report measures of prejudice do at least possess face validity: we feel we know what it means when respondents say they like one group more than another or endorse stereotypes about groups. Thus, in a sense, the verbal behavior validates the underlying attitude (Fazio, 2007), rendering further evidence of behavior prediction unnecessary to the understanding of explicitly endorsed prejudice. For implicitly measured prejudice, however, predictive validity is crucial because instantiations of the implicit prejudice construct lack face validity as measures of intergroup *prejudice*. Indirect measures of prejudice lack face validity because these measures avoid having respondents consciously endorse malevolent or benevolent forms of

prejudice. Accordingly, linking scores on implicit prejudice measures to behavior is crucial to show that the tests tap into a psychological construct that affects how groups are perceived and treated beyond the narrow confines of the implicit test. If whatever it is that is measured by the IAT or another implicit measure reliably predicts behaviors that can contribute to intergroup conflict, then concerns about the lack of definitional and theoretical clarity about underlying processes and inputs would be reduced (but, even from an applied perspective, one should still care about theory because a causal account may be needed to formulate policy aimed at reducing bias and preventing discrimination).

All measures of implicit prejudice employ indirect approaches, and many are reaction-time-based measures for which millisecond differences in response times may be taken as evidence of a prejudicial attitude. It would hardly be surprising if fans of the Chicago Cubs associated the New York Yankees more quickly with success than the Chicago Cubs, or more quickly identified the word "winner" as being a positive term after seeing a picture of a Yankee than a picture of a Cub. But, are shorter latencies in response times sufficient to declare the Cub fan (implicitly) prejudiced against the Cubs, especially when different implicit measures produce different results and when different word pairings on the same type of measure may produce different associations and different results? Most laypersons, as well as many scholars, understand prejudice (whether preceded by the modifier "implicit" or not) to extend beyond mere negative or positive associations with an attitude object to include affective and motivational reactions to in-group and out-group members (e.g., Allport, 1954; Brown, 1995; Duckitt, 2003).

One might take the extreme nominalist position that implicit prejudice need refer to nothing more than reaction time differences on a measure of implicit bias (analogous to the old positivist view that IQ is whatever IQ tests measure). But the many uses of the implicit prejudice construct outside academic psychology that treat implicit prejudice as having motivational and behavioral implications indicate that the public does not understand implicit prejudice in this nominalist way. If one circularly defines an implicit attitude to equal one's score on an implicit measure, and if scores on these measures fail to predict any judgments or behaviors reliably, then the concept of implicit prejudice is meaningless, except in the context of measurement. In that case, one could make implicit bias go away simply by stopping use of the implicit measure.

Perhaps most tellingly, defenders of the implicit prejudice construct often revert to claiming that measures of implicit prejudice predict discriminatory behavior as the justification for treating implicit prejudice as a type of prejudice (e.g., Banaji, Deutsch, & Banse, 2004; Banaji & Greenwald, 2013; Gawronski et al., 2011; Greenwald et al., 2009; Nosek & Greenwald, 2009). Or, as one reviewer of *Blindspot* wrote when discussing the IAT as a measure of prejudice: "*The best indicator of the test's validity is its prediction of behavior*" (Hutson, 2013, emphasis added). And, with *Blindspot* as his guide, this reviewer concluded that the IAT does predict discriminatory behavior, and does so "even better than do overt statements about one's beliefs" (Hutson, 2013).

But does the IAT really do better than explicit measures at predicting discrimination? Given the statements made in *Blindspot* and elsewhere about the supposed predictive superiority of the IAT over self-report measures, it may be surprising to learn that the answer is "no." Greenwald and his colleagues (2009) reported that, for most IATs included in their meta-analysis of criterion studies, *explicit measures outperformed IATs in the prediction of judgments, decisions, and behavior in seven of the nine criterion domains studied.* By Greenwald et al.'s (2009) own numbers, explicit measures outperformed IATs even in a number of domains where social desirability bias should have been at work with respect to the explicit measures, including interactions with the other gender and with persons of different sexual orientations, alcohol and drug use, and psychological health. Although Greenwald et al. (2009) wrote that "for socially sensitive topics, the predictive power of self-report measures was remarkably low and the incremental validity of IAT measures was relatively high" (Greenwald et al., 2009, p. 32), in actuality only race IATs and a collection of IATs lumped under the heading "other intergroup behavior" (which included weight, age and ethnicity IATs) outperformed explicit measures, and this superior performance was primarily due to the poor performance of the explicit measures in the race and "other intergroup behavior" domains. In fact, the race and other-intergroup IATs synthesized in Greenwald et al.'s meta-analysis performed at or below the predictive validity found for explicit measures of prejudice in other meta-analyses that have examined the relation between explicit prejudice and behavior ($r = 0.24$ and 0.20, respectively, for the race and other intergroup IATs in Greenwald et al., 2009, versus $r = 0.26$ in Kraus, 1995, and $r = 0.24$ in Talaska et al., 2008, for explicit measures of prejudice).[7]

However, there are good reasons not to rely on Greenwald et al.'s (2009) estimates of predictive validity for even the race and "other intergroup" IATs. First, Greenwald and colleagues (2009) utilized a meta-analytic approach that aggregated across many different conditions and masked the degree of variability present in the studies. For instance, Greenwald et al. treated brain wave activity while watching black and white faces on par with micro-level behaviors in interracial interactions, which were treated as on par with explicit judgments and choices toward white and black persons (i.e., type of criterion measure was not treated as a moderator variable). Second, the moderator variables that Greenwald et al. did examine were confounded with criterion domain, and no within-domain moderators were reported. Third, Greenwald et al. failed to include a number of effects (e.g., only the effects for behavior directed at blacks but not at whites were included for a number of the synthesized studies).

To address the shortcomings in Greenwald et al. (2009), we (and colleagues) conducted an updated and expanded meta-analysis of studies in which scores from race or ethnicity IATs were correlated with criterion measures (Oswald et al., 2013). Using this expanded database, we found substantially lower estimates of predictive validity for the IATs than those reported by Greenwald et al. (2009). (Recently, Nosek and colleagues conducted a meta-analysis estimating the correlation between behavior and implicit bias as measured by the IAT or any other implicit measure in

an experimental setting, and they found a mean correlation even lower than we found; see Forscher et al., 2016). We also found that explicit measures of prejudice performed at approximately the same, and sometimes slightly higher, levels than IATs, and this result held whether the criterion variable involved micro-level or macro-level behavior. Indeed, in studies using response times on a task as the criterion variable, explicit measures were more predictive than IATs. This result casts into doubt theories of implicit attitude–behavior relations (and corresponding public statements such as those found in *Blink*) in which implicit bias is portrayed as more predictive of spontaneous, subtle behaviors than deliberate behaviors. Consistent with the poor predictive validity of both implicit and explicit measures of prejudice, we found that the measures alone or together explained small amounts of variance in behavior, with neither adding much incremental validity to the other measure. Furthermore, we found tremendous variance in results across studies. In many instances, the variance was much greater than the estimated effect size. Thus, regardless of where one's score on the IAT places one under Project Implicit's bias classification system (test-takers are told they have no automatic preference for one group over another, a slight automatic preference, a moderate automatic preference, or a strong automatic preference), one's score on the IAT will be a poor predictor of whether one will act fairly or unfairly toward a minority group member. In a positive sign for the power of data to influence the implicit prejudice dialogue, Greenwald, Banaji, and Nosek (2015) recently agreed with this conclusion, although debate continues over whether the small effects observed for implicit bias within aggregated data may accumulate over time to produce societal harms (see Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Based on the existing research, it would be a high-risk gamble to predict even aggregate patterns of behavior of any kind from IAT scores, and one would fare just as well, and often better, at the betting table by basing one's bets on scores from explicit measures of prejudice than on IAT scores.

Were one to read only popularizations of IAT research, one would conclude that the IAT is a better predictor of discriminatory behavior than explicit measures of prejudice, and in particular subtle and spontaneous forms of discrimination. And one might conclude that this is true whether the discrimination takes the form of in-group favoritism or out-group antagonism. Both of those conclusions would be false.

The score interpretation problem

If criterion studies do not provide the basis for characterizing particular IAT scores as indicative of no, low, moderate, or high bias, then on what basis are visitors to Project Implicit given feedback about their level of personal bias, and on what basis is 75% of the American public being described by IAT researchers as implicitly racist?[8] It turns out that test-taker feedback and the distribution of implicit racism provided by the IAT's creators are based on arbitrary and shifting judgments that have nothing to do with external validation of the meaning of IAT scores. The IAT's

creators simply adapted to their test Cohen's effect size rule of thumb for gauging the size of a psychological effect and established cut-points for different levels of bias without public explanation of those cut-points (see Blanton & Jaccard, 2008), even though Cohen made clear that his rule of thumb was arbitrary and that effect sizes should be linked to practical measures of meaning and significance. This behaviorally untethered approach to score interpretation means that, should the IAT's creators change their mind about bias cut-points, then individual test-takers could be given very different feedback, and the prevalence of implicit bias could be shifted by researcher fiat. In fact, such a shift has already occurred once. As Blanton and Jaccard (2008) discuss, in connection with the replacement of the original IAT scoring algorithm, the IAT's creators also changed their criteria for categorizing the extremity of an IAT score, and, as a result, the percentage of persons supposedly showing strong anti-black bias on the IAT dropped from 48% to 27%. This change in levels of implicit prejudice was not due to a sudden societal shift, nor due to the findings of any studies linking particular bands of IAT scores to particular behaviors. This change was due solely to the researchers' change in definitions.

This degree of researcher freedom to make important societal statements about the level of implicit prejudice in American society, with no requirement that those statements be externally validated through some connection to behavior or outcomes, points to the potential mischief that attends a test such as the IAT that employs an arbitrary metric. Unlike scales for physical quantities and (some) explicit measures of prejudice, which have intuitively meaningful zero points and gradations in the scales, scores on the IAT (which are transformed difference scores that are the product of algorithmic calculations) have no clear meaning without a supplemental context.[9] To say that one person has a higher score on the IAT than another does not mean that the former person is more likely to express bias in some way outside the testing context. Given that the existing correlational data on the bias–behavior relation is weak and highly unreliable (as discussed earlier), there is no empirical justification presently even for taking a dichotomous approach to IAT scores, under which particularly high scores would be treated as evidence of bias and scores below that threshold would not.[10]

The relativistic nature of the IAT, in which one attitude category is contrasted with another, compounds the problem, for persons with different associations and association strengths for the opposing attitude objects may receive similar IAT scores (e.g., a person who associates the category "European Americans" a bit more quickly with positive terms than the category "African-Americans" may receive the same score as someone who associates the category "African-Americans" a bit more quickly with negative terms than the category "European Americans"). Absent evidence linking difference scores on the IAT to observable behaviors, and absent evidence showing that persons in the same bias categories reliably show the same behavioral patterns, it is impossible to give meaning and practical significance to IAT scores.

When Project Implicit tells a test-taker that his IAT score reveals a strong automatic preference for whites, all that really means is that the test-taker's relative

reaction times, as measured in milliseconds, were above a threshold arbitrarily set by the test designers (i.e., the bias label is just shorthand for reaction time differences – it is not shorthand for bias on anything other than the test). This score, and the bias category assigned to it, do not have any behavioral significance beyond the test itself, yet statements on Project Implicit strongly imply that they do have behavioral meaning for the individual.[11]

The implicit sexism puzzle

One particularly puzzling aspect of academic and public dialogue about implicit prejudice research has been the dearth of attention paid to the finding that men usually do not exhibit implicit sexism, while women do show pro-female implicit attitudes (e.g., Lemm & Banaji, 1999; Nosek, 2005; Nosek & Banaji, 2001; Skowronski & Lawrence, 2011). These findings are contrary to the common finding on IATs of the historically advantaged group being favored by members of both the advantaged and disadvantaged groups. Because of the continuing importance of male–female disparities in occupational representation and wages, and because of ongoing problems of female victimization (Rudman & Mescher, 2012), the implicit prejudice meme is often extended to the problem of gender discrimination (e.g., Sandberg, 2013; Vedantam, 2010). The implicit sexism findings thus present both theoretical and societal puzzles to be solved, yet these findings have received little attention.

The response to the null findings of implicit sexism among men has been to focus on findings showing that men are more commonly associated with math and science and with a limited set of leadership qualities. Thus, if one visits Project Implicit, one will likely find an IAT aimed at assessing whether men or women are more quickly associated with science or math terms (a gender stereotype IATs), but one will probably not find an IAT aimed at assessing implicit attitudes toward men and women (perhaps due to the robustness of the null implicit-sexism finding with respect to men). But one will also not find on Project Implicit a discussion of how men typically do not exhibit implicit sexism.

The focus on implicit gender stereotypes, and away from implicit sexism, is problematic for practical and theoretical reasons. First, implicit measures of gender stereotypes are not good predictors of discriminatory behavior. Recall that Greenwald et al. (2009) found that explicit measures of gender prejudice outperformed gender IATs in predicting behavior, and neither type of measure explained even 5% of the variance in behavior (a finding consistent with many other studies of gender bias effect sizes). Small effect sizes may have noticeable practical effects over time or in large samples under some circumstances, and of course individual instances of sex discrimination occur, but the implicit gender-bias studies provide no basis for expecting measures of implicit gender stereotypes or implicit sexism to be reliable predictors of practically significant adverse effects on particular employment decisions (see, e.g., Derous, Ryan, & Serlie, 2015).

Second, only a very limited set of implicit gender stereotypes has been examined. For instance, research has not examined whether many traits of good managers, such as cooperativeness, fairness, and integrity, are more strongly associated with women than men. There is no reason to believe that only the few gender stereotypes examined to date are the only stereotypes that may be activated consciously or subconsciously, and there is no reason to believe that implicit gender stereotypes are universally negative for women for all positions for which they compete with men.

Third, no explanation is provided for how conflicts between automatic evaluative associations and automatic semantic associations are resolved, but the gap between implicit attitudes and stereotypes raises important theoretical questions about the linkage of evaluative and semantic associations. If the forms that prejudice and discrimination take depend on how feelings and beliefs interact (e.g., Kervyn, Fiske, & Yzerbyt, 2013), then an account of this interaction is needed at the implicit level just as it is at the explicit level. That is, something more than opportunistic citations to the set of findings that support the implicit prejudice meme is needed. An account is needed for why we should expect an implicit negative stereotype to have a powerful negative influence on behavior but a positive implicit attitude to have no influence on behavior, and we need a contextualized model of when and where each component of implicit prejudice will be predominant, and in what behavioral form. This question poses a problem that cuts across many situations where group identities intersect, and in which the associations with some identities or features of identities are positive and some are negative. As we discuss in the next section, given the nature of implicit prejudices, we would posit that, in our increasingly multicultural world, many situations present cases of intersecting identities, and thus the potential for many conflicting and reinforcing biases (see Hewstone, Turner, Kenworthy, & Crisp, 2006).

If bias = (relative) association strength, then bias is everywhere

Banaji and Greenwald define implicit attitudes as nothing more or less than evaluative associations of varying strengths with attitude objects, whether those objects be products, places, or people, and whether the source of those associations be cultural information or personal experience (e.g., Banaji & Greenwald, 2013; Banaji & Heiphetz, 2010; Banaji et al., 2004; cf. Fazio, 2007). Under this view of attitudes, implicit prejudice is just a type of evaluative knowledge about different groups, with some evaluations being more positive and some more negative than others. When two groups for which one holds evaluative associations are placed in opposition (as they are on the IAT, given that it assesses only relative reaction times), then one may be said to be implicitly prejudiced in favor of one group or against the other group whenever associations with the groups are not in the same direction and are not approximately equal in strength (with strength operationalized as relative reaction times on the IAT).

If any source of evaluative associations can be the source of implicit attitudes, and if implicit bias means a non-random difference in the relative strength of evaluative associations as measured by the IAT, then implicit biases should be rampant. Under the logic of the IAT, a wide variety of biases beyond those based on the traditional legally protected categories of race, ethnicity, gender, age, disability, and religion should be identifiable, and in fact a good number of nontraditional biases have been identified (e.g., Democrat vs. Republican, liberal vs. conservative, married vs. single, Northerners vs. Southerners, rich vs. poor) (see Nosek, 2005). But we have only scratched the surface of implicit biases, given the expansive view of implicit prejudice that underlies the implicit prejudice meme. Some research has examined the impact of competing implicit biases on intersecting categories of traditional concern, such as how Dutch women fare relative to Muslim males (Derous et al., 2015), but, to our knowledge, no research has sought to examine the many nontraditional implicit biases that may be implicated in an interaction and compare the behavioral influence of the nontraditional biases to that of the traditional implicit biases.

One unexamined explanation for the weak correlations found between implicit measures and criterion variables (Greenwald et al., 2009; Oswald et al., 2013) is that a welter of unmeasured implicit biases create tremendous noise and counteracting effects in any given situation where people activate multiple categories and evaluative/semantic associations. Moreover, in the face of this welter of group-based implicit knowledge, associations at a more localized level, such as implicit associations with a particular individual formed through personal interactions, may have precedence over more general associations, as work by Quinn and Macrae indicates (Quinn & Macrae, 2005; Quinn, Mason, & Macrae, 2009). In other words, just as individuating information exerts powerful effects that counter explicit biases, it does the same with respect to implicit biases.

The subjective judgment problem

Whenever moderators of the bias–behavior relation are discussed, the situational factors commonly invoked as enabling the expression of implicit bias are subjectivity in judgment and discretion in decision-making (e.g., Hart, 2005; Heilman & Haynes, 2008). Likewise, one common strategy offered to prevent the influence of implicit bias is to objectify judgment and decision-making processes to the greatest extent possible (e.g., the recommendation to rely only on objective measures of performance for employee assessments). These contentions derive almost entirely from experimental studies in which (a) subjects who are inexperienced with performing certain tasks are given discretion in how to judge (b) hypothetical persons or strangers about whom they have very limited information. These articles rarely acknowledge the many experimental studies in which subjective judgment is not associated with the expression of bias (see, e.g., Swim, Borgida, & Maruyama, 1989). Also, more importantly, articles positing subjectivity as the doorway to implicit-bias-based discrimination never deal with the large amount of research

from industrial–organizational studies finding that subjective evaluation criteria are not associated with discrimination against women and minorities in real organizations (e.g., Hennessey & Bernardin, 2003; MacKay & McDaniel, 2006; Roth, Huffcutt, & Bobko, 2003). These findings cannot be reconciled with the implicit prejudice meme: if implicit prejudice is not evident in subjective employment judgments and decisions, then it is not plausible to assume that it will be evident in more objective judgments and decisions, nor is it clear how the micro-level aggressions that implicit prejudice is posited to produce are leading to adverse employment outcomes. One can think of subjective performance evaluations as presenting a not-very-stringent test of the implicit prejudice meme, but the meme fails even this test.

What are the real contributions?

Reading *Blindspot* (Banaji & Greenwald, 2013), one is struck by how little of the information presented there originated with the IAT or even with implicit prejudice research more generally. That is not a criticism of the book, which is aimed at presenting a picture of social cognition as often affected by automatic processes that can have detrimental and surprising consequences. Arguably, the only significant message that derives uniquely from the IAT work is that humans are beset by many implicit biases, often at surprisingly high rates of prevalence. However, given that the bias categories associated with the IAT have not been externally validated, and given that an individual IAT score is itself only a moderately reliable predictor of future IAT scores, the social significance of the widespread biases identified by the IAT is unclear. And, given that considerable confusion remains about the nature of implicit prejudice and its links to behavior despite the considerable resources and attention devoted to IAT research – indeed, the implicit prejudice construct is arguably even more contested among social psychologists now than it was before the IAT era – the theoretical contributions of IAT research are also unclear.

In terms of practical contributions, one could argue that the assistance of plaintiffs in litigation through expert witness services is a practical contribution, but that conclusion depends, in our view, on the validity of the claims made by the witnesses. If one looks for effective diversity or anti-discrimination programs that are based on IAT research, one will look in vain, for few bias-reduction techniques have proven behaviorally potent in experimental settings (see Forscher et al., 2016; Lai et al., 2014), and none have been shown to reduce discrimination or increase diversity in a real-world setting. Indeed, in a *Wall Street Journal* article on the increasing popularity of incorporating research on the IAT into diversity training, Greenwald expressed skepticism about its utility: "Professor Greenwald warns that 'unconscious-bias training often is just window dressing' that fails to alter work practices" (Lublin, 2014).

By developing a test that reliably produces statistically significant results, and by making it easy for individual researchers to use and adapt the IAT for their own purposes, the IAT's creators have produced a tool that is nearly self-perpetuating: as more researchers publish results based on the tool, the greater the collective

motivation to justify use of the tool and its outputs. Whether this tool will have a longer lifespan than many other popular tools and research paradigms in social psychology (for examples, see Greenwald, 2012) remains to be seen. Regardless of the length of that life, there is no doubt that the IAT energized the study of prejudice among social psychologists and brought to this field of inquiry many who might not otherwise have entered it.

Why does the implicit prejudice meme persist?

The ease with which the IAT can be used to produce statistically significant effects, along with the possibility that these effects reveal subterranean biases that might account for widespread societal inequalities, offer a ready explanation for the initial appeal of the IAT among psychologists and for the rise of the IAT-inspired implicit prejudice meme. The persistence of the meme in the face of accumulating evidence of conceptual confusion, psychometric uncertainties, and predictive disappointments is harder to explain. Our best guess is that the continuing popularity of the IAT, and the determination to read social significance into the pattern of aggregate IAT data, reflects a confluence of ideological sympathies, publication bias, and the lack of clear, consensual score-keeping measures within social psychology.

The liberal bias and the bias in favor of publishing non-null, experimental results among the editors of psychology journals are hard to dispute (e.g., Gross, 2013; Inbar & Lammers, 2012; Mitchell, 2012). Thus, it is not surprising that IAT studies are easy to publish, or that the IAT has attracted the interest of many socially conscious psychologists.

Theoretical battles within social psychology are not so much won as endured until boredom and exhaustion set in (Meeehl, 1967; Tetlock & Manstead, 1985) – a state of affairs that reflects how many methodological and theoretical degrees of freedom sparring partners have to elude stringent empirical tests of the sort found in the physical sciences, with no obligation to produce something of demonstrated practical value (Meehl, 1967, 1978, 1990; Tetlock, Mellers, Rohrbaugh, & Chen, 2014). The focus of the IAT on unconscious structures and processes that are not directly observable and the loose tethering of operationalizations of the prejudice and discrimination constructs to the real-world events that the constructs are meant to explain provide IAT researchers and IAT-research translators with many degrees of interpretive freedom, both to expand the explanatory scope of the implicit prejudice construct and to deflect challenges to the implicit prejudice meme. Thus, we find Dr. Banaji, in her keynote address at a recent Association for Psychological Science convention, moving seamlessly between the IAT as reflecting associative learning versus reflecting the degree to which one identifies with different social groups, and we find her claiming that patterns of IAT results reflect system justification tendencies (Jaffe, 2014). No doubt, Banaji can invoke operationalizations of "identification" and "system justification" to support her claims, and, more importantly, she will be able to dispute evidence that supposedly conflicts with her claims

on grounds that improper operationalizations were employed. The implicit prejudice meme has at its disposal willing and capable defenders operating in a space of seemingly endless protective moves. Add to the mix the ideological sympathy that many social psychologists likely have for the implicit prejudice meme and the hurdles that a meme skeptic will face in terms of funding and publication, and the road to reduce the popularity of the IAT through ordinary science looks long and difficult to navigate.

Given the potential payoff from better understanding the causes of inequality, and given the real expenditures being made on IAT research and in response to fears caused by the implicit prejudice meme, the implicit prejudice research domain is fertile ground for experimentation with extraordinary science. We have discussed at length what one form of extraordinary science might look like in this domain (Tetlock & Mitchell, 2009), but that is only one possibility. What is essential is that ground rules for judging success and failure be set *ex ante* rather than allow contestants to engage in *post hoc* assimilation of any pattern of results to their preferred theories. Rather than wait on the slow evolution of scientific knowledge, an ambitious funding agency should finance an empirical tournament requiring transparency in predictions, methods, data, and results, which requires researchers to declare *ex ante* their priors and to state how surprising different results would be, and which imposes external, objective measures of success (Tetlock & Mitchell, 2009; Tetlock et al., 2014).

Absent an embrace of extraordinary science along these lines, we suspect that the implicit prejudice meme will persist outside academia so long as the implicit prejudice construct remains more an idea than a guide to practical solutions. For once employers, health care providers, police forces, and policy-makers seek to develop real solutions to real problems and then monitor the costs and benefits of these proposed solutions, the shortcomings of implicit prejudice research will likely become apparent outside of academia.

Conclusion

Just as social psychologists were puzzling over how the decline in explicit prejudice could be reconciled with ongoing inequalities and seeking to develop psychology-based answers (e.g., Dovidio & Gaertner, 2000), the IAT arrived on the scene. The IAT was not the first implicit measure of prejudice, but it was the first measure supposedly to reveal pervasive implicit biases against a wide range of historically disadvantaged groups, and to do so reliably (at least in the aggregate). Excited by the IAT results, the IAT's creators and a host of allies took to the public airwaves to broadcast these results, to describe the IAT as revolutionary, and, most importantly, to extrapolate from the IAT results to a wide range of social relations. This extrapolation has seemingly known no bounds, including the bounds of empirical science, for many of the public claims made by the IAT's boosters have little empirical support, and a number of those claims are counter to the existing empirical record. If scientific

success were measured only by citation counts and number of mentions in public discourse, then the IAT would be a resounding success. However, if scientific success is measured by the degree to which behavior in real-world settings can be explained and predicted, then the IAT falls far short. The IAT is too useful a rhetorical tool to be discarded on merely scientific grounds. Legal-political actors can use the test to make aggressive claims about the pervasiveness and potency of bias in any policy arena of their choosing. But, as William Blake noted in his Proverbs of Hell, we often find out we have had enough only after we have had more than enough. There was value to warning society about the dangers of under-estimating bias, but there is also value to warning society of the dangers of over-estimating bias, of using wobbly science to support far-reaching claims.

Endnotes

1 Another important early step in the growth of the popularity of the IAT was the decision by its creators to support development of the Inquisit software for implementation of the test (see http://www.millisecond.com/about/about.aspx), and to share with other researchers test stimuli and the code for analyzing IAT data (see http://faculty.washington.edu/agg/iat_materials.htm and http://projectimplicit.net/nosek/iat). Within just a few years of the IAT's introduction, hundreds of IAT studies had been published. IATs now exist for self-assessments (e.g., self-esteem and risk of self-injury), for product assessments (e.g., Coke vs. Pepsi), for assessing implicit attitudes toward various behaviors and activities (e.g., smoking and drug use), and, of course, for assessing implicit prejudice against a wide variety of groups (e.g., prejudicial attitudes and stereotypes with respect to elderly persons, women, Muslims, and persons with disabilities).

2 Gladwell and Greenwald appeared in the following year on *The Oprah Winfrey Show* to discuss the IAT (http://www.oprah.com/oprahshow/Overcoming-Prejudice).

3 The Project Implicit website tells potential customers that "[i]mplicit measures have a variety of potential applications such as market research, organizational behavior, health and medicine, human factors, law, public policy, and judgment and decision-making. Many clients will collaborate with Project Implicit to conduct research, or contract with Project Implicit to implement and host novel applications of implicit measures for research, education, or organizational purposes" (http://projectimplicit.net/customwebsites.html).

4 For instance, in his discussion about the IAT on the Edge.org website, Greenwald downplayed criticisms of the IAT: "The test has critics, but of about 500 scientific publications on the IAT so far, perhaps two or three percent are critical" (http://www.edge.org/conversation/the-implicit-association-test).

5 They add that "[i]mportantly, implicit biases in one's thoughts are known to affect one's decisions, actions, and judgments, producing discriminatory effects whether or not they were consciously intended by the decision-maker" (Dasgutpa & Stout, 2012, p. 400).

6 Aversive racism researchers posit processes other than association strength as drivers of aversive racism, such as motivated shifting of evaluative standards and differential weighing of evidence (Hodson, Dovidio, & Gaertner, 2002). Gawronski and colleagues (e.g., Brochu, Gawronski, & Esses, 2008; Gawronski, Peters, Brochu, & Strack, 2008) treat

aversive racism as a higher-level type of prejudice that describes a system of processes and inputs, including implicit bias.

7 Cameron et al. (2012), in their meta-analysis of the predictive validity of sequential priming measures, also reported that the priming and explicit measures performed comparably.

8 The "Frequently Asked Questions" page of Project Implicit contains the following question and answer: "What does it mean that my IAT score is labeled 'slight', 'moderate', or 'strong'? If you respond faster when flower pictures and pleasant words are paired on a single key than when insect pictures and pleasant words are paired on a single key, we would say that you have an implicit preference for flowers relative to insects. The labels slight, moderate and strong reflect the strength of the implicit preference – how much faster do you respond to flowers + pleasant versus insects + pleasant" (https://implicit.harvard.edu/implicit/faqs.html#faq3).

9 Not all explicit measures of prejudice have face validity. Self-report scales aimed at measuring modern or new forms of prejudice have engendered debate over the meaning and implications of their scores (see Biernat & Crandall, 1999).

10 As Uhlmann and colleagues discuss, it is not appropriate to base individual diagnostic assessments on correlational data. An individual's IAT score "is not independently informative about the individual. Rather, this value is only meaningful in the context of a greater data set, and only for prediction. ... Thus, researchers should be careful in the conclusions they draw and the recommendations they make from the use of these measures. Future research using implicit measures should aim to develop norms and cut-off scores, the usual way of creating nonarbitrary metrics in psychological and managerial research" (Uhlmann et al., 2012, p. 582).

11 If one is troubled by one's score on the IAT, and the feedback that accompanies it, then a good way to reduce one's bias (on the test) is to take the test again. Because the test has only moderate test–retest reliability, and because the test is subject to practice effects that result in performance on IAT blocks converging (Nosek, Greenwald & Banaji, 2007), subsequent scores are likely to reveal evidence of reduced bias. Thus, one approach to eliminating implicit prejudice as measured by the IAT, and as reported to the public by the IAT's creators, would be to have all Americans repeatedly take the IAT until they show no bias on the test.

References

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

Andreychik, M. R., & Gill, M. J. (2012). Do negative implicit associations indicate negative attitudes? Social explanations moderate whether ostensible "negative" associations are prejudice-based or empathy-based. *Journal of Experimental Social Psychology, 48*, 1082–1093.

Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Asosciation Test?" *Psychological Inquiry, 15*, 257–278.

Babcock, P. (2006). Detecting hidden bias. *HR Magazine, 51*. Available at http://www.shrm.org/publications/hrmagazine/editorialcontent/pages/0206cover.aspx)

Bagenstos, S. R. (2007). Implicit bias, "science," and antidiscrimination law. *Harvard Law and Policy Review, 1*, 477–493.

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

Banaji, M. R. (2008, August). The science of satire. *The Chronicle Review, 54*(31), B13.

Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In D. T. Gilbert & S. T. Fiske (Eds.), *Handbook of social psychology* (Vol. 1, pp. 353–393). Hoboken, NJ: John Wiley & Sons.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68*, 181–198.

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people.* New York, NY: Random House.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes & Tetlock. *Psychological Inquiry, 15*, 279–289.

Bar-Anan, Y., & Nosek, B. A. (2012). *A comparative investigation of seven implicit measures of social cognition.* Unpublished manuscript. University of Virginia.

Benforado, A., & Hanson, J. (2008). Legal academic backlash: The response of legal theorists to situationist insights. *Emory Law Journal, 57*, 1087–1145.

Biernat, M., & Crandall, C. S. (1999). Racial attitudes. In J. P. Robinson, P. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 297–411). New York, NY: Academic Press.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology, 34*, 277–297.

Blasi, G., & Jost, J. T. (2006). System justification theory and research: Implications for law, legal advocacy, and social justice. *California Law Review, 94*, 1119–1168.

Blow, C. M. (2009, February 20). A nation of cowards? *New York Times.*

Brochu, P. M., Gawronski, B., & Esses, V. M. (2008). Cognitive consistency and the relation between implicit and explicit prejudice: Reconceptualizing old-fashioned, modern, and aversive prejudice. In M. A. Morrison & T. G. Morrison (Eds.), *The psychology of modern prejudice* (pp. 27–50). Hauppauge, NY: Nova Science Publishers.

Brown, R. (1995). *Prejudice: Its social psychology* (1st edn). Oxford, UK: Blackwell Publishing, Ltd.

Brown, R. (2010). *Prejudice: Its social psychology* (2nd edn). Oxford, UK: Blackwell Publishing, Ltd.

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

Chapman, E. N., Kaatz, A., & Carnes, M. (2013). Physicians and implicit bias: How doctors may unwittingly perpetuate health care disparities. *Journal of General Internal Medicine, 28*, 1504–1510.

Chugh, D. (2004). Societal and managerial implications of implicit social cognition: Why milliseconds matter. *Social Justice Research, 17*, 203–222.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin, 87*, 546–563.

Crouch, M. A., & Schwartzman, L. H. (2012). Introduction. *Journal of Social Philosophy, 43*, 205–211.

Dasgupta, N., & Stout, J. G. (2012). Contemporary discrimination in the lab and real world: Benefits and obstacles of full-cycle social psychology. *Journal of Social Issues, 68*, 399–412.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwartz (Eds.), *Implicit measures of attitudes* (pp. 179–194). New York, NY: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347–368.

Derous, E., Ryan, A. M., & Serlie, A. W. (2015). Double jeopardy upon resumé screening: When Achmed is less employable than Aïsha. *Personnel Psychology, 68*, 659–696.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*, 315–319.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 1084–1121). New York, NY: Wiley.

Duckitt, J. (2003). Prejudice and intergroup hostility. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 559–600). Oxford, UK: Oxford University Press.

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology, 54*, 297–327.

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.

Feingold, J., & Lorang, K. (2012). Defusing implicit bias. *UCLA Law Review Discourse, 59*, 210–228.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2016). *A meta-analysis of change in implicit bias*. Unpublished manuscript.

Garda, R. A. Jr. (2011). The white interest in school integration. *Florida Law Review, 63*, 599–655.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology, 50*, 141–150.

Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 78–123). New York, NY: Guilford Press.

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

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

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34*, 648–665.

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York, NY: Little, Brown and Company.

Gomez, M. R. (2013). The next generation of disparate treatment: A merger of law and social science. *Review of Litigation, 32*, 553–589.

Goode, E. (1998, October 13). A computer diagnosis of prejudice. *New York Times*, F7.

Gove, T. G. (2011). Implicit bias and law enforcement. *The Police Chief, 78*, 44–56. Available at http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=print_display&article_id=2499&issue_id=102011)

Green, T. K. (2010). Race and sex in organizing work: "Diversity," discrimination, and integration. *Emory Law Journal, 59*, 585–647.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science, 7*, 99–108.

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

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*, 945–967.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65–82). Hillsdale, NJ: Erlbaum.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. *Meta-analysis of predictive validity. Journal of Personality and Social Psychology, 97*, 17–41.

Gross, N. (2013). *Why are professors liberal and why do conservatives care?* Boston, MA: Harvard University Press.

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

Han, H. A., 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.

Hardin, C. D., & Banaji, M. R. (2012). The nature of implicit prejudice: Implications for personal and public policy. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 13–31). Princeton, NJ: Princeton University Press.

Hart, M. (2005). Subjective decisionmaking and unconscious discrimination. *Alabama Law Review, 56*, 741–791.

Heilman, M. E., & Haynes, M. C. (2008). Subjectivity in the appraisal process: A facilitator of gender bias in work settings. In E. Borgida & S. T. Fiske (Eds.), Beyond common sense: Psychological science in court (pp. 127–155). Oxford, UK: Blackwell Publishing, Ltd.

Hennessey, H. W. Jr., & Bernardin, H. J. (2003). The relationship between performance appraisal criterion specificity and statistical evidence of discrimination. *Human Resource Management, 42*, 143–158.

Hewstone, M., Turner, R. N., Kenworthy, J. B., & Crisp, R. J. (2006). Multiple social categorization: Integrative themes and future research priorities. In R. J. Crisp & M. Hewstone (Eds.), *Multiple social categorization: Processes, models and applications* (pp. 271–310). New York, NY: Psychology Press.

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin, 28*, 460–471.

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2004). The aversive form of racism. In J. L. Chin (Ed.), *The psychology of prejudice and discrimination* (Vol. 1, pp. 119–135). Westport, CT: Praeger.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

Hutson, M. (2013, February 8). Review of *Blindspot*. *Washington Post*.

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7*, 496–503.

Jaffe, E. (2014, July/August). The science of "us" and "them." *APS Observer, 27*, 7–8.

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

Kang, J. (2012). Communications law: Bits of bias. In J. D. Levinson & R. J. Smith (Eds.), *Implicit racial bias across the law* (pp. 132–145). Cambridge, MA: Cambridge University Press.

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of "affirmative action." *California Law Review, 94*, 1063–1118.

Kang, J., Bennett, M., Carbado, D., Casey, P., Dasgupta, N., Faigman, D., Godsil, R., Greenwald, A., Levinson, J., & Mnookin, J. (2012). Implicit bias in the courtroom. *UCLA Law Review, 59*, 1124–1186.

Kervyn, N., Fiske, S. T., & Yzerbyt, Y. (2013). Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity). *European Journal of Social Psychology, 7*, 673–681.

Kester, J. D. (2001, July/August). A revolution in social psychology. *APS Observer Online, 14*. Available at http://www.psychologicalscience.org/observer/0701/family.html.

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21*, 58–75.

Kristof, N. D. (2008, April 6). Our racist, sexist selves. *New York Times*. Available at http://www.nytimes.com/2008/04/06/opinion/06kristof.html.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., Sartori, G., Dial, C., Sriram, N., Banaji, M. R., & Nosek, B. A. (2014). A comparative investigation of 17 interventions to reduce implicit racial preferences. *Journal of Experimental Psychology: General, 143*, 1765–1785.

Lane, K. A., Kang, J., & Banaji, M. R. (2007). Implicit social cognition and law. *Annual Review of Law and Social Science, 3*, 427–451.

Lemm, K., & Banaji, M. R. (1999). Unconscious beliefs and attitudes about women and men. In U. Pasero & F. Braun (Eds.), *Wahrnehmung und Herstellung von Geschlecht* (pp. 215–233). Opladen: Westdutscher Verlag.

Levinson, J. D., & Smith, R. J. (Eds.) (2012). *Implicit racial bias across the law*. Cambridge, MA: Cambridge University Press.

Lieber, L. D. (2009). The hidden dangers of implicit bias in the workplace. *Employment Relations Today, 36*, 93–98.

Lublin, J. S. (2014, January 9). Bringing hidden biases into the light – big businesses teach staffers how "unconscious bias" impacts decisions. *Wall Street Journal*. Available at http://online.wsj.com/news/articles/SB10001424052702303754404579308562690896

Mahajan, N., Martinez, M., Gutierrez, N. L., Diesendruck, G., Banaji, M., & Santos, L. R. (2011). The evolution of intergroup bias: Perceptions and attitudes in rhesus macaques. *Journal of Personality and Social Psychology, 100*, 387–405.

Mahajan, N., Martinez, M., Gutierrez, N. L., Diesendruck, G., Banaji, M., & Santos, L. R. (2014). Retraction of Mahajan, Martinez, Gutierrez, Diesendruck, Banaji, & Santos (2011). *Journal of Personality and Social Psychology, 106*, 182.

McKay, P. F., & McDaniel, M. A. (2006). A reexamination of black–white mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91*, 538–554.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103–115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108–141.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*, 109–117.

Mitchell, G., & Tetlock, P. E. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal, 67*, 1023–1121.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625–666.

Nosek, B. A., & Greenwald, A. G. (2009). (Part of) the case for a pragmatic approach to validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135*, 373–376.

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

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

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

Paul, A. M. (1998, May 1). Where bias begins: The truth about stereotypes. *Psychology Today, 31*.

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the affect misattribution procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin, 39*, 375–386.

Potier, B. (2004, December 16). Making case for concept of "implicit prejudice": Extending the legal definition of discrimination, *Harvard University Gazette*. Available at http://www.news.harvard.edu/gazette/2004/12.16/09-prejudice.html.

Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology, 32*, 299–328.

Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology, 88*, 467–479.

Quinn, K. A., Mason, M. F., & Macrae, C. N. (2009). Familiarity and person construal: Individuating knowledge moderates the automaticity of category activation. *European Journal of Social Psychology, 39*, 852–861.

Reeves, A. R. (2012, May 1). Diversity in practice: The power of a hoodie. *Chicago Lawyer.* Available at http://www.chicagolawyermagazine.com/Elements/pages/print.aspx?printpath=/Archives/2012/05/20691&classname=tera.gn3article

Richardson, L. S. (2011). Arrest efficiency and the Fourth Amendment. *Minnesota Law Review, 95,* 2035–2098.

Robinson, R. K. (2008). Perceptual segregation. *Columbia Law Review, 108,* 1093–1180.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88,* 694–706.

Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and male sexual aggression. *Personality and Social Psychology Bulletin, 38,* 734–746.

Sandberg, S. (2013). *Lean in: Women, work, and the will to lead.* New York, NY: Knopf.

Shermer, M. (2006, November 24). Comic's outburst reflects humanity's sin. *L.A. Times.* Retrieved from Westlaw Newsroom, 2006 WLNR 20385662.

Siegel, E., Dougherty, M. R., & Huber, D. E (2012). Manipulating the role of cognitive control while taking the implicit association test. *Journal of Experimental Social Psychology, 48,* 1057–1068.

Siegel, E., Sigall, H., & Huber, D. E. (2012). The IAT is sensitive to the perceived accuracy of newly learned associations. *European Journal of Social Psychology, 42,* 189–199.

Skowronski, J. J., & Lawrence, M. A. (2011). A comparative study of the implicit and explicit gender attitudes of children and college students. *Psychology of Women Quarterly, 25,* 155–165.

Swim, J., Borgida, E., & Maruyama, G. (1989). Joan McKay vs. John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105,* 409–429.

Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimating racial discrimination: A meta-analysis of the racial attitude–behavior literature shows that emotions, not beliefs, best predict discrimination. *Social Justice Research, 21,* 263–296.

Tetlock, P. E., Mellers, B. A., Rohrbaugh, N., & Chen, E. (2014). Forecasting tournaments: Tools for increasing transparency and improving the quality of debate. *Perspectives on Psychological Science, 23,* 290–295.

Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? In B. M. Staw & A. Brief (Eds.), *Research in organizational behavior* (Vol. 29, pp. 3–38). New York, NY: Elsevier.

Tibbits, G. (1998, October 12). Prejudice test. *Associated Press Online.*

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods, 15,* 553–601.

Vedantam, S. (2010). *The hidden brain: How our unconscious minds elect presidents, control markets, wage wars, and save our lives.* New York, NY: Random House Publishing Groups.

Wilson, T., Lindsey, S. & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101–126.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.

Yzerbyt, V., & Demoulin, S. (2010). Intergroup relations. In S. T. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 1024–1083). New York, NY: Wiley.