



Review

# 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

John T. Jost <sup>a,\*</sup>, Laurie A. Rudman <sup>b</sup>, Irene V. Blair <sup>c</sup>, Dana R. Carney <sup>d</sup>,  
Nilanjana Dasgupta <sup>e</sup>, Jack Glaser <sup>f</sup>, Curtis D. Hardin <sup>g</sup>

<sup>a</sup> *New York University, Department of Psychology, 6 Washington Place, 5th Floor, New York, NY 10003, United States*

<sup>b</sup> *Rutgers University, United States*

<sup>c</sup> *University of Colorado, Boulder, United States*

<sup>d</sup> *Columbia Business School, United States*

<sup>e</sup> *University of Massachusetts, Amherst, United States*

<sup>f</sup> *Goldman School of Public Policy, University of California at Berkeley, United States*

<sup>g</sup> *Brooklyn College and Graduate Center, City University of New York, United States*

Available online 8 November 2009

## Abstract

In this article, we respond at length to recent critiques of research on implicit bias, especially studies using the Implicit Association Test (IAT). Tetlock and Mitchell (2009) claim that “there is no evidence that the IAT reliably predicts class-wide discrimination on tangible outcomes in any setting,” accuse their colleagues of violating “the injunction to separate factual from value judgments,” adhering blindly to a “statist interventionist” ideology, and of conducting a witch-hunt against implicit racists, sexists, and others. These and other charges are specious. Far from making “extraordinary claims” that “require extraordinary evidence,” researchers have identified the existence and consequences of implicit bias through well-established methods based upon principles of cognitive psychology that have been developed in nearly a century’s worth of work. We challenge the blanket skepticism and organizational complacency advocated by Tetlock and Mitchell and summarize 10 recent studies that no manager (or managerial researcher) should ignore. These studies reveal that students, nurses, doctors, police officers, employment recruiters, and many others exhibit implicit biases with respect to race, ethnicity, nationality, gender, social status, and other distinctions. Furthermore—and contrary to the emphatic assertions of the critics—participants’ implicit associations do predict socially and organizationally significant behaviors, including employment, medical, and voting decisions made by working adults.

© 2009 Elsevier Ltd. All rights reserved.

## Contents

1. Thirty years (or more) of research on implicit bias: a primer . . . . .	42
1.1. Cognitive accessibility as a method of demonstrating implicit bias . . . . .	43
1.2. Semantic priming as a method of demonstrating implicit bias . . . . .	43
1.3. Evaluative priming as a method of demonstrating (and measuring) implicit bias . . . . .	44
1.4. Individual differences in the motivation to control prejudice . . . . .	44

\* Corresponding author.

*E-mail address:* [john.jost@nyu.edu](mailto:john.jost@nyu.edu) (J.T. Jost).

1.5.	Enter the Implicit Association Test . . . . .	45
1.6.	The upshot . . . . .	46
2.	Ten studies of implicit bias that no manager should ignore . . . . .	46
2.1.	Managerial decisions concerning callback interviews . . . . .	49
2.2.	Hiring preferences, budget cuts, verbal slurs, social exclusion, and physical harm . . . . .	49
2.3.	Police officers' decisions to shoot. . . . .	50
2.4.	Medical decisions concerning patient treatment and job turnover . . . . .	51
2.5.	Undecided voters: for whom will they (eventually) vote? . . . . .	52
2.6.	Binge drinking, pedophilia, and suicide attempts . . . . .	52
3.	Other problems with the Tetlock–Mitchell critique. . . . .	53
3.1.	A refutation of philosophical and ideological objections . . . . .	53
3.1.1.	Have you heard the news? . . . . .	54
3.1.2.	Facts have nothing to do with it. . . . .	55
3.1.3.	Caricaturing the goals and assumptions of implicit bias research . . . . .	56
3.1.4.	The parable of the “Two Jesses” (once more with feeling) . . . . .	57
3.2.	A refutation of methodological objections . . . . .	59
3.2.1.	Are implicit biases due to “benign” causes? . . . . .	59
3.2.2.	Familiarity. . . . .	59
3.2.3.	Cultural awareness . . . . .	60
3.2.4.	Sympathy for the disadvantaged. . . . .	61
3.2.5.	Fear of being labeled a bigot, not bigotry . . . . .	61
3.2.6.	Cognitive dexterity . . . . .	62
3.3.	Is organizational accountability a panacea? . . . . .	62
4.	Concluding remarks . . . . .	63
	Acknowledgement . . . . .	64
	References . . . . .	64

---

*Prejudice is not “the invention of liberal intellectuals.” It is simply an aspect of mental life that can be studied as objectively as any other. (Allport, 1954, p. 516)*

The job of the social scientist, one could say, is to investigate those characteristics of society that ordinary citizens are unable, unwilling, or perhaps too uninterested to observe, analyze, and understand. Some citizens may even resist the notion that the subject matter taken for granted by social and behavioral scientists—such as the idiographic origins of our attitudes or the shared nature of our prejudices—is amenable to “expert” interpretation. Taking the long view of it, it seems likely that open societies benefit when certain individuals take it upon themselves to investigate both the “wildly abnormal and the tediously normal” aspects of human behavior, as the philosopher MacIntyre (1958, p. 1) described the sundry manifestations of the unconscious. In the short run, however, it is easy to understand how some would respond with annoyance if not alarm at the snooping, probing, and seemingly audacious allegations of the social scientists in their midst. Resistance is all the more likely when social scientific discoveries seem to challenge long-cherished personal or cultural assumptions, such as the relatively hopeful messages that (a) human thought and behavior are largely under the control of individual will and consciousness, and (b) racial prejudice in Western societies (especially the U.S.) is a thing of the past.

The discovery of implicit or automatic bias by social and cognitive psychologists over the past 30 years has rankled much of the laity and certain pockets of academic and political life on both of these scores. But it is at least conceivable that the discomfort that some feel when first confronted with the possibility that they could be harboring biases of which they have little or no awareness is the mark of the discovery (e.g., see Fiske, 2003). From this perspective, the most rancorous reactions would be expected to hound those social scientific insights that underscore the least praiseworthy inclinations of ordinary folks, for such insights are the ones that can touch a nerve. It is worth recalling, for instance, that Myrdal, author of *An American Dilemma* (1944)—a social scientific treatise on racial prejudice that was cited approvingly in the Supreme Court’s *Brown v. Board of Education* ruling that ended desegregation in the schools—was denounced by a Senator from Mississippi and in numerous Southern newspapers as a member of the “international Communist conspiracy” (Herman, 1995). Given the historical context, it was probably inevitable that research on implicit prejudice would attract a number of scholarly as well as ideological objections. But reading the

critique by Tetlock and Mitchell (2009)—or any of the other highly similar, predominantly overlapping critiques of research on implicit prejudice that Tetlock has previously co-authored (Arkes & Tetlock, 2004; Mitchell & Tetlock, 2006; Tetlock & Arkes, 2004; Tetlock & Mitchell, 2008)<sup>1</sup>—one is likely to come away with a sense that the discipline of social psychology itself is in disarray and that its constituents are deeply divided over the question of whether implicit bias really exists. *This impression would be false.* The fact of the matter is that the field of social psychology has not seriously doubted the existence of implicit stereotyping and prejudice since at least the early 1990s, if not earlier.

We synopsise the history of several decades of work in social and cognitive psychology to make clear that the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which is a reaction-time measure that captures the strength with which social groups (and other attitude objects) are implicitly or automatically associated with good/bad evaluations and other characteristics, is merely one methodological innovation in a long stretch of empirical contributions documenting the existence of implicit bias. In some ways, it should not be surprising that a technique for measuring implicit prejudice that is as pervasive and popular as the IAT<sup>2</sup> would attract negative as well as positive attention. By singling out IAT researchers for special criticism, however, Tetlock and Mitchell (2009) obscure the extent to which their reservations (as skeptical outsiders without any formal training in the use of implicit methods) are in conflict not only with a handful of individual researchers but with several decades' worth of solid research conducted by scores of prominent experimental psychologists.

Skepticism plays a critical role in scientific progress, and the best, most constructive of the critiques have led to significant improvements in the scoring and administration of the IAT, as the test's designers have themselves acknowledged (e.g., see Greenwald, Nosek, & Banaji, 2003; Lane, Banaji, Nosek, & Greenwald, 2007; Nosek, Greenwald, & Banaji, 2007). At the same time, however, a minority of critics, including Tetlock and Mitchell (2009), leave the unfortunate impression that researchers of implicit bias are up to something distasteful, presumptuous, and even ethically objectionable. In a law review article that covers much of the same ground as the current critique, they wrote that:

It is easy to be overwhelmed by the sheer volume of laboratory studies that implicit prejudice advocates cite, by the moral certitude with which they apply psychological generalizations to the real world, and by the impressive credentials they bring to the courtroom. But this would be a big mistake. On closer inspection, we shall discover that the scientific rhetoric accompanying legal applications of this research is . . . more honorific than descriptive. (Mitchell & Tetlock, 2006, p. 1029)

Mitchell and Tetlock (2006) also accuse researchers of implicit bias of violating a “canonical scientific norm,” namely “the injunction to separate factual from value judgments.”<sup>3</sup> They write:

Attributions of prejudice inevitably rest on complex amalgams of factual and value assumptions, and it is a mistake to suppose that, just because a select group of social psychologists and law professors—with a self-

<sup>1</sup> It is troubling to some implicit bias researchers that few, if any, of these critiques appear to have been published as a result of the peer review process that is a hallmark of the scientific enterprise and that serves as a much-needed mechanism of quality control. The critics will surely counter that those under attack are hardly able to serve as “unbiased” reviewers. However, because social and cognitive psychologists have been studying implicit attitudes for well over 30 years, there are plenty of qualified reviewers who are outside of the relatively small circle of researchers that has been targeted by Tetlock and Mitchell.

<sup>2</sup> A PsycINFO search for “Implicit Association Test” in the title, abstract, and keywords yielded over 490 publications. Because the measure was introduced in 1998, this reflects an average of almost 50 articles per year. Millions of participants around the world have now taken the IAT. Its broader cultural popularity is reflected in the fact that it has been featured on ABC's *Nightline* and *20/20*, in *Scientific American* and numerous other print and Internet media sources, in popular non-fiction such as Malcolm Gladwell's *Blink*, and even the cartoon situation comedy *King of the Hill*.

<sup>3</sup> Tetlock has been accusing colleagues, especially prejudice researchers, of violating this “canonical scientific norm” for over twenty years now (e.g., Sniderman and Tetlock, 1986; Tetlock, 1994). Curiously, Tetlock and Mitchell (2009) believe that “on topics as divisive as racial inequality, it is impossible to insulate science from political disputes” (p. 6, emphasis added; see also Tetlock, 1994). This means that they are criticizing prejudice researchers for failing to do that which, according to their own view, is impossible. The response offered by Sears (1994), one of Tetlock's perennial targets, provides a useful counterpoint to the injunction to remain silent on questions of moral and political value, which seems to be the only way out of the “Catch-22” situation set up by Tetlock and Mitchell. Sears suggested that spelling out one's theoretical and political positions is “far healthier than cloaking our own feelings in a pretense of scientific objectivity, while ignoring a ream of scientific evidence we happen to find distasteful” (p. 555). Indeed, it is difficult to imagine the history of prejudice research in social science without any discussion of moral and political values. Who among us would retract the contributions made by Gunnar Myrdal, Gordon Allport, or Kenneth Clark or their impact on public discourse concerning race relations or civil rights legislation?

declared agenda to transform American law—announce the discovery of a new form of prejudice, the rest of society is obliged to defer to their judgment. (Mitchell & Tetlock, 2006, p. 1032)

Without claiming that anyone is “obliged” to defer to any particular judgment, we wish merely to register our dissent from Tetlock and Mitchell’s *de facto* decree to “stay out of it.” As another prejudice researcher, Allport (1941), once wrote in urging his colleagues to become *more* rather than less involved in public affairs: “If the psychologist is tempted to say that he knows too little about the subject he may gain confidence by watching the inept way in which politicians, journalists, and men in public life fence with the problems of propaganda, public opinion, and morale” (p. 235).

To be sure, individual researchers can be mistaken. However, the history of modern science suggests that, though there may be plenty of room for misconceptions at the margins, it is rare (though not entirely unheard of) for the huge preponderance of scientists within a well-established discipline to be utterly and spectacularly wrong. This seems to be what Tetlock and Mitchell (2009) are suggesting—that the near-consensus among social psychologists concerning the discovery of implicit bias reflects some kind of collective folly or perhaps a hoax or in any case the regrettable result of a blinding ideological commitment to the doctrine of “statist interventionism”<sup>4</sup> (see also Wax & Tetlock, 2005). The fact of the matter is that, as we will endeavor to illustrate, the existence of implicit bias is beyond reasonable doubt. There are not “islands of consensus in a sea of controversy,” as Tetlock and Arkes (2004) claimed; rather, it is very much the other way around.

In this article, we aim to convey our intellectual excitement over the most important achievements that have emerged from the last 30 years’ worth of social psychological research on implicit bias as well as its ramifications for our evolving understanding of the nature of prejudice. In so doing, we challenge the blanket skepticism and organizational complacency advocated by Tetlock and Mitchell (2009) concerning the existence of implicit bias. Following a brief historical primer of scientific developments leading up to the discovery of implicit bias, we focus on 10 recent studies that vividly illustrate the power of implicit racial, gender, and other biases to affect both the judgments and behaviors of students, nurses, doctors, police officers, and employment recruiters, and other mere mortals. Although—as Tetlock and Mitchell (2009) contend—organizational decision-makers such as business managers might be considered exempt from such human shortcomings, this has not turned out to be the case with respect to other biases and heuristics, and it strikes us as such a highly risky bet that it violates good sense for executives (and their attorneys) to place it.<sup>5</sup> In short, Tetlock and Mitchell are *dispensing bad advice when they counsel organizational researchers, practitioners, and business leaders to ignore the problem of implicit bias*. The evidence, including the research we summarize here, is strong enough to conclude that implicit bias is a genuine phenomenon that requires thoughtful analysis and a meaningful consideration of practical interventions and forms of redress.<sup>6</sup>

## 1. Thirty years (or more) of research on implicit bias: a primer

Adapting well-established scientific methods and principles from the field of cognitive psychology, a cohort of leading researchers—including Banaji and Greenwald (1995), Banaji, Hardin, and Rothman (1993), Bargh and Pratto (1986), Bodenhausen (1990), Darley and Gross (1983), Devine (1989), Dovidio, Evans, and Tyler (1986), Dovidio, Kawakami, Johnson, Johnson, and Howard (1997), Fazio, Jackson, Dunton, and Williams (1995), Fazio, Sanbonmatsu, Powell, and Kardes (1986), Gaertner and McLaughlin (1983), Macrae, Bodenhausen, Milne, and Jetten (1994), Perdue and Gurtman (1990), Rudman and Borgida (1995), and Stangor, Sullivan, and Ford (1991), among many others—provided a solid empirical bedrock for understanding the occurrence of implicit bias. These studies, which have been replicated and extended in literally hundreds of peer-reviewed articles, established that people can

<sup>4</sup> Tetlock and Mitchell (2009) use the term “statist interventionism” to characterize ideological support—allegedly on the part of political liberals—for high levels of governmental (or state) intrusion into the free market in order to address problems such as employment discrimination.

<sup>5</sup> We agree with Tetlock and Mitchell that it would be extremely useful to investigate implicit bias in the workplace directly. If cooperative organizations could be found, it would indeed present a valuable opportunity for collaboration (adversarial or otherwise)—to confirm that implicit bias affects the decisions made by members of “real-world” organizations, as Rooth (2007) has found (see also Green et al., 2007).

<sup>6</sup> For other qualitative and quantitative reviews of research programs on implicit bias (see Dasgupta, 2004; Greenwald et al., 2009; Hardin and Banaji, *in press*; Rudman, 2004). For discussions of the societal and legal implications of these findings (see Blasi and Jost, 2006; Faigman, Dasgupta, & Ridgeway, 2008; Greenwald and Krieger, 2006; Jolls and Sunstein, 2006; Kang, 2005; Kang and Banaji, 2006; Krieger and Fiske, 2006; Mentovich and Jost, 2008; but see Mitchell and Tetlock, 2006; Wax, 2008, for dissenting views).

possess attitudes, stereotypes, and prejudices in the absence of intention, awareness, deliberation, or effort (cf. Bargh, 1994).

The power of expectations (based on learned contingencies) to unwittingly affect basic perception, judgment, memory, and behavior had already been established in experiments that mark the advent of empirical psychology (e.g., Bartlett, 1932; Ebbinghaus, 1885/1913; Hebb, 1949; Stroop, 1935). But it was research on automatic semantic associative links in memory—mental connections among words and concepts that are meaningfully related to one another (Meyer & Schvaneveldt, 1971, 1976; Neely, 1977)—that led most directly to the discovery that stereotyping and prejudice might also operate implicitly. By the early-1980s, the scientific consensus regarding the importance and ubiquity of nonconscious, highly efficient, and automatized cognition was firmly established (e.g., Ratcliff & McKoon, 1994, 1988; Shiffrin & Schneider, 1977; see also Hasher & Zacks, 1979). The notion that judgment and decision-making (of experts and layperson alike) is subject to unintentional biases and shortcomings was also central to the monumental contributions summarized by Kahneman, Slovic, and Tversky (1982) and Nisbett and Ross (1980). These research programs, in turn, built on Simon's (1957) concept of “bounded rationality” and related work illustrating the limitations of effortful information processing.

The first demonstrations of implicit stereotyping and prejudice, then, were merely logical extensions of a well-known cognitive principle—namely that knowledge is organized in memory in the form of semantic associations that are derived from personal experiences as well as normative procedures and rules. Thus, it may be news to some that implicit prejudice is an empirical possibility, but the truth is that its discovery followed naturally from a century's worth of research on perception, memory, and learning. The phenomenon of implicit bias was comfortably assimilated into theories of mundane, workaday principles of human information processing, and it fits the contemporary consensus in the brain and behavioral sciences that an enormous amount of cognition occurs automatically, effortlessly, and outside of conscious awareness (e.g., Bargh, 1994; French & Cleeremans, 2002).

### *1.1. Cognitive accessibility as a method of demonstrating implicit bias*

Implicit bias was demonstrated initially in two complementary experimental paradigms that emerged from the social information-processing nexus of cognitive and social psychology. The first of these was developed by Devine (1989) in a watershed article that adapted experimental procedures in which social perception and memory were shown to be influenced by exposure to semantically (or stereotypically) related information (e.g., Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979; see also Darley & Gross, 1983). In Devine's critical experiment, participants evaluated “Donald” as more hostile after they had been subliminally exposed to a relatively large (vs. small) proportion of words related to common stereotypes of African Americans in a previous, ostensibly unrelated task. Not only were participants' social judgments affected by stereotypes unintentionally and in the absence of awareness once they had been rendered accessible through subliminal exposure, but this effect on social judgment occurred to a similar degree regardless of how participants had scored on explicit measures, that is, when they had voluntarily reported their personal racial attitudes. This basic effect, which demonstrates that the cognitive salience of a familiar stereotype can implicitly bias social judgment in stereotype-consistent ways, was replicated dozens of times with respect to a wide variety of social groups around the world (see Devine & Monteith, 1999; Devine, 2005; Dovidio & Gaertner, 2004; Fiske, 1998). Conceptual replications include but are not limited to studies of implicit attitudes concerning ethnicity and race (Devine, 1989; Rudman & Lee, 2002), gender (Banaji & Greenwald, 1995; Banaji et al., 1993; Rudman & Borgida, 1995), and age (Gross & Hardin, 2007; Levy, 1996).

### *1.2. Semantic priming as a method of demonstrating implicit bias*

A second paradigm adapted cognitive measures of semantic association (e.g., Neely, 1977) to investigate implicit prejudice. According to this paradigm, social attitudes—including prejudice and stereotypes—are empirically captured by the degree to which they are linked through speed and efficiency to semantically related concepts. For example, exposure to a word that is either denotatively or connotatively related to women (e.g., lady, nurse) hastens the speed with which people identify female pronouns that appear subsequently; similarly, exposure to a word linked to men (e.g., gentleman, doctor) facilitates the identification of male pronouns (Banaji & Hardin, 1996; Blair & Banaji, 1996).

In hundreds of experiments involving thousands of participants and a plethora of specific experimental variations—including serial semantic priming techniques (Blair & Banaji, 1996; Fazio et al., 1995), eye-blink startle responses

(Amodio, Harmon-Jones, & Devine, 2003), and cognitive/behavioral interference paradigms such as Stroop tasks (Bargh & Pratto, 1986; Richeson & Trawalter, 2005) and IATs (Greenwald et al., 1998; Greenwald, Nosek, et al., 2003)—research has demonstrated the prevalence of stereotypical and prejudicial associations. Target groups have again been wide-ranging, including those based on ethnicity and race (Jost, Banaji, & Nosek, 2004; Nosek, Banaji, & Greenwald, 2002), gender (Banaji & Hardin, 1996), sexual orientation (Dasgupta & Rivera, 2006; Jost et al., 2004), body shape (e.g., Bessenoff & Sherman, 2000), the elderly (Perdue & Gurtman, 1990), adolescents (Gross & Hardin, 2007), and even fictitious groups created in the laboratory (DeSteno, Dasgupta, Bartlett, & Caidric, 2004).

### 1.3. *Evaluative priming as a method of demonstrating (and measuring) implicit bias*

In one of the earliest compelling demonstrations of implicit intergroup bias, Fazio et al. (1995) employed a computer-based task known as evaluative (or sequential) priming. Specifically, they exposed participants to photographs of either white or black faces and asked them to categorize subsequent words as either positive or negative, as quickly as possible. Each experimental trial (of which there were hundreds) was composed of a single face prime and a single target word to which the response was made. It is important to note that the presentation speed of the faces and words was short enough in duration to ensure that their influence on responses would be the result of automatic rather than controlled processes, at least according to 30 years' worth of research (e.g., Neely, 1977; see also Fazio et al., 1986).

Results indicated that when white participants classified positively valenced words, their responses were faster when they had been exposed to white (vs. black) faces, but when they classified negatively valenced words, their responses were faster when they had been exposed to black (vs. white) faces.<sup>7</sup> Fazio et al.'s (1995) research program was among the first to investigate systematic differences in reaction times as a function of race and valence as an unobtrusive measure of intergroup attitudes. The pattern of findings has been widely interpreted by social psychologists as indicating the presence of implicit racial bias in favor of whites relative to blacks (but see Arkes & Tetlock, 2004, for a different view).

Fazio et al. (1995, Study 1) also included a procedure in which a black assistant, who was unaware of participants' implicit and explicit bias scores, interacted with the participants and rated them on their levels of interest and friendliness toward her. Participants' scores on the Modern Racism Scale did not predict the quality of the interaction as rated by their assistant, but their implicit bias scores did ( $r = .31, p < .05$ ). This finding provided early evidence that implicit attitudes, as indexed by a carefully constructed measure, can predict the quality of cross-racial interactions. Subsequent research using independent raters viewing videotaped interactions has confirmed this initial result (e.g., Amodio & Devine, 2006; Dovidio et al., 1997; Dovidio, Kawakami, & Gaertner, 2002; Richeson & Shelton, 2005). Towles-Schwen and Fazio (2006) conducted a longitudinal study of white and black students randomly assigned to be college roommates. At the beginning of the year, they assessed whites' degree of implicit prejudice. At the end of the year, they found that 28% of cross-race relationships had ended (compared with only 3% for same-race roommates), and that whites' implicit bias scores predicted shorter durations of their cross-race relationships,  $r(55) = .30, p < .05$ .

### 1.4. *Individual differences in the motivation to control prejudice*

Fazio et al. (1995) also identified an important moderator of the relationship between implicit and explicit biases, namely the "motivation to control prejudice." Specifically, they developed a questionnaire measure that included items such as, "It's never acceptable to express one's prejudices." Not surprisingly, participants who were higher in motivation to control prejudice scored lower on an explicit measure of prejudice (the Modern Racism Scale). By contrast, the motivation to control prejudice did not predict implicit bias, but it did moderate the relation between implicit and explicit forms of bias. Specifically, implicit bias was more strongly correlated with explicit bias when the motivation to control prejudice was low than when it was high—an observation that is consistent with recent meta-analytic findings that the correlation between implicit and explicit attitudes is relatively weak with respect to socially

<sup>7</sup> In Fazio et al.'s (1995) research, black participants tended to show the opposite pattern, apparently reflecting implicit ingroup favoritism in general. However, studies using other experimental paradigms have found that members of disadvantaged groups (including African Americans) show less implicit ingroup favoritism and sometimes even outgroup favoritism (e.g., see Ashburn-Nardo et al., 2003; Jost et al., 2004; Livingston, 2002; Nosek et al., 2002; Rudman et al., 2002).

sensitive attitudes such as race but strong when it comes to attitudes people are comfortable sharing in public (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

This research and the many studies on implicit and explicit motivations to control prejudice that followed (Amodio et al., 2003; Amodio, Kubota, Harmon-Jones, & Devine, 2006; Amodio, Devine, & Harmon-Jones, 2008; Dasgupta & Rivera, 2006; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Dunton & Fazio, 1997; Glaser & Knowles, 2008; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999; Park, Glaser, & Knowles, 2008; Plant & Devine, 1998; Son Hing, Li, & Zanna, 2002), make abundantly clear that the goal of research on implicit bias is certainly not to conclude that “We are all racists at heart,” as Wax and Tetlock (2005) hyperbolically alleged (more on this later). Although research suggests that anyone who is familiar with a stereotype is at least capable of drawing on it (intentionally or unintentionally), there are indeed meaningful differences in the magnitude of expressed implicit bias as a function of individual differences in the motivation to be unbiased (which can be conscious but need not be), as the citations above attest. There are also interpersonal and structural features of the social situation that moderate the expression of implicit bias (e.g., Lowery, Hardin, & Sinclair, 2001; Sinclair, Lowery, Hardin, & Colangelo, 2005).

As this review of research on cognitive accessibility, semantic priming, evaluative priming, and other methods and principles of implicit social cognition demonstrate, literally hundreds of studies provide conclusive evidence that mental processes (including mental processes about individuals and groups) can and do operate nonconsciously and can be measured implicitly. Why should anyone—more than 50 years after Allport (1954) got things started—think that mental processes involved in stereotyping and prejudice would be any different? To take Tetlock and Mitchell’s critique seriously, one would need to set aside so much of social and cognitive psychology that both disciplines would be rendered unrecognizable to contemporary students and scholars.

### 1.5. Enter the Implicit Association Test

Fazio et al.’s (1995) research paved the way for the IAT (Greenwald et al., 1998; Greenwald, Nosek, et al., 2003), which is now the most widely used method for measuring implicit bias. The IAT also uses response latencies to measure implicit attitudes and follows a similar logic as evaluative priming (and, indeed, the famous “Stroop Task,” which preceded both techniques by several decades; Stroop, 1935). The IAT gauges differences in how easy or difficult it is for people to associate individual exemplars of various social categories (whites vs. blacks, rich vs. poor, gay vs. straight, and so on) with abstract words and categories that have evaluative implications (e.g., good vs. bad, pleasant vs. unpleasant). Thus, people who are faster to categorize the faces or names of whites when they are paired with positive (vs. negative) stimuli and, conversely, the faces or names of blacks when they are paired with negative (vs. positive) stimuli, are theorized to have internalized a stronger preference for whites relative to blacks, compared to people who respond more equivalently across different category-valence pairings (or in the opposite direction).<sup>8</sup> Nosek and Smyth (2007) have summarized evidence that the IAT exhibits construct, convergent, and divergent validity and have also discussed its relationship to more traditional (i.e., explicit self-report) measures of attitudes (see also Nosek, Smyth, et al., 2007).<sup>9</sup>

Phelps et al. (2000) even provided physiological evidence for the construct validity of implicit attitudes in general and the predictive validity of the IAT in particular when they demonstrated that IAT scores were correlated with the magnitude of amygdala activation when white participants were exposed to photographs of unfamiliar black (vs. white) faces. This is important because the amygdala is involved in emotional responses to threat, such as fear (LeDoux, 1996). The fact that self-reported racial attitudes *failed* to predict amygdala activation under the same circumstances strengthens the already burgeoning case for the discriminant validity of implicit attitudes. Mendes, Gray, Mendoza-Denton, Major, and Epel (2007) provide further evidence that the IAT predicts psychophysiological reactions in interracial encounters. Specifically, they found that whites who exhibited stronger implicit racial bias reacted with greater physiological stress when speaking to a black audience. The important point to distill from these findings is that the IAT shows good construct validity through its association with neural systems serving as reliable indicators of individual differences (see also Amodio et al., 2003; Mitchell, Macrae, & Banaji, 2006). None of this evidence for the validity of the IAT is acknowledged (let alone explained) by the critics.

<sup>8</sup> Readers can experience the race IAT (and many other tests) for themselves by visiting Project Implicit (<https://implicit.harvard.edu/implicit/>).

<sup>9</sup> Detailed information about the validity of the IAT is also available at [http://faculty.washington.edu/agg/iat\\_validity.htm](http://faculty.washington.edu/agg/iat_validity.htm).

### 1.6. *The upshot*

Tetlock and Mitchell (2009) characterize the major suppositions of implicit bias research as “extraordinary claims” that “require extraordinary evidence” (p. 11). Although we do see some of the evidence, including the physiological evidence, as extraordinary in terms of its specificity and perspicacity, our historical primer should help to clarify that researchers of implicit bias have not been making “extraordinary claims,” at least if one accepts the major tenets of 20th century cognitive psychology. We suspect that for most readers, the evidence we have summarized already would be enough to suggest that implicit bias is a genuine, observable phenomenon that deserves to be taken seriously. In the remainder of this article, we will review a number of additional studies—including several that are quite recent and overlooked (or at least uncited by) the skeptics. Many of these recent studies speak more directly than their predecessors to the specific concerns about predictive validity raised by Arkes and Tetlock (2004), Mitchell and Tetlock (2006), and Tetlock and Mitchell (2008, 2009). At this juncture, we wish to emphasize that the case for implicit bias in no way depends upon any single methodological innovation (such as the IAT), nor is it restricted to associations about race or ethnicity or gender. The variety of methods used to study implicit attitudes about a wide range of stimuli—including Stroop and lexical decision tasks, supraliminal and subliminal priming procedures, and linguistic bias coding techniques—are rooted in several decades’ worth of research in cognitive and social psychology and are rigorously empirical; they should not be equated with “mind-reading,” as Mitchell and Tetlock (2006) suggested.

In order for Tetlock and Mitchell’s (2009) deeply skeptical argument to succeed, they would need to debunk not merely the same handful of IAT studies that Arkes and Tetlock (2004), Mitchell and Tetlock (2006), Blanton, Jaccard, Gonzales, and Christie (2006), and Wax (2008) have repeatedly picked on and picked over (i.e., the studies by McConnell & Leibold, 2001; see response by McConnell & Leibold, 2009; Ziegert & Hanges, 2005). Rather, Tetlock and Mitchell would need to discount or reject nearly a 100 years of cognitive psychology and more than three decades of paradigm-shifting research from dozens of the most well respected social psychological laboratories in the world. This is a burden they have not come close to meeting in this or any of their earlier critiques (see Arkes & Tetlock, 2004; Mitchell & Tetlock, 2006; Tetlock & Arkes, 2004; Tetlock & Mitchell, 2008; Wax & Tetlock, 2005). Managerial and organizational researchers and practitioners should not be misled; rather, they deserve to know that the existence of implicit bias is widely accepted by the scientific community of social psychologists, and with good reason.

## 2. Ten studies of implicit bias that no manager should ignore

The Tetlock–Mitchell case for organizational complacency with respect to implicit bias rests precariously on their blunt assertion that “there is no evidence that the IAT reliably predicts class-wide discrimination on tangible outcomes in any setting” (p. 6). This claim, which we regard as specious, also provides the cornerstone for earlier, equally dismissive critiques of research on implicit bias (e.g., Mitchell & Tetlock, 2006; Tetlock & Mitchell, 2008; Wax, 2008). Here again, Tetlock and Mitchell (2009) are misleading organizational researchers and practitioners by implying that the evidence for the predictive validity (or behavioral significance) of implicit bias is weak and can be easily dismissed. The truth is that such evidence is already strong, and it continues to grow in depth and breadth.

We already know that Mitchell and Tetlock (2006) are unimpressed “by the sheer volume of laboratory studies that implicit prejudice advocates cite” (p. 1029). Nevertheless, we recommend that interested readers consult several qualitative and quantitative (i.e., meta-analytic) reviews that comprehensively document clear, consistent, and impressive (especially by the standards of behavioral science) correlations between (a) measures of implicit attitudes and explicitly held judgments and evaluations (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005), and (b) measures of implicit attitudes and consequential behaviors, including prejudicial and discriminatory behaviors (Dasgupta, 2004; Greenwald et al., 2009). The material they review systematically refutes Tetlock and Mitchell’s claim that “there is no evidence that the IAT reliably predicts class-wide discrimination” (p. 6). Rather than duplicating these efforts, we focus instead on a relatively small subset of recent studies that have different methodological strengths but, when taken in conjunction, provide a compelling empirical case for the predictive validity of implicit bias measures, including the IAT. These studies should not be ignored by anyone who is responsible for hiring, firing, interviewing, managing, or evaluating others (see Table 1 for an executive summary of these studies).



Table 1  
Executive summary of 10 studies that no manager should ignore.

Brief summary (1–2 sentences)	Source/citation	Measure of implicit bias	Behavioral outcome	Research background/elaboration/replications/extensions
(1) Employment recruiters who favored native Swedes over Arabs on an implicit stereotyping task were significantly less likely to offer (equally qualified) Arab (vs. Swedish) applicants job interview opportunities. Overall, Swedes were 3 times more likely to receive callback interviews.	Rooth (2007)	IAT (evaluative stereotyping)	Contacting job applicants for interviews	Bertrand and Mullainathan (2004) found that job applicants with African American names were 50% less likely to receive job interviews than those with European American names (see also Bertrand et al., 2005).
(2) Despite the fact that participants regarded female (vs. male) managerial applicants who presented themselves as confident, competitive, and ambitious as highly qualified, they also disliked them and were therefore less likely to recommend hiring them. Participants' implicit gender stereotyping predicted the extent of disliking.	Rudman and Glick (2001)	IAT (gender stereotyping)	Liking ratings (mediator of hiring bias)	Much research suggests that the stereotypical prescription for female "niceness," which operates at implicit and explicit levels of awareness, can penalize women in the workplace (e.g., Eagly & Karau, 2002; Fiske et al., 1991; Heilman et al., 2004).
(3) White student participants who scored higher on measures of implicit bias against various racial/ethnic outgroups were significantly more likely to report engaging in verbal slurs, social exclusion, and physical harm against members of minority groups. They also were more likely to recommend budget cuts disproportionately against Jewish, Asian, and black student associations.	Rudman and Ashmore (2007)	IAT (evaluative stereotyping)	Verbal slurs, social exclusion, physical harm, budget cuts	Implicit attitudes were found to possess strong incremental validity, predicting harmful and discriminatory behavior even after adjusting for explicit racial and ethnic attitudes.
(4) In the context of a video simulation program, police officers were significantly more likely to "shoot" an unarmed suspect when he was black (vs. white) on early trials, but they were able to overcome this bias with practice.	Plant and Peruche (2005)	Split-second (simulated) shooting decisions of black vs. white suspects		Eberhardt et al. (2004) found that police officers possessed implicit associations linking black faces to criminality. Correll et al. (2007) showed that police officers and civilians were (a) quicker to "shoot" armed black (vs. white) suspects, and (b) slower to refrain from shooting unarmed black (vs. white) suspects. Research by Glaser and Knowles (2008) revealed that people who implicitly associated blacks with weapons also tended to shoot armed blacks faster than armed whites in a video simulation.
(5) Physicians' degree of implicit (but not explicit) bias predicted racial disparities in simulated treatment recommendations. Specifically, greater bias was associated with (a) decreased likelihood of recommending thrombolysis for black patients suffering from coronary heart disease, and (b) increased likelihood of recommending it for comparable white patients.	Green et al. (2007)	IAT (racial preferences)	Simulated treatment recommendations for heart patients	For years, medical researchers have sought to determine why white patients with symptoms of myocardial infarction are nearly twice as likely as blacks with similar symptoms to receive thrombolytic therapy (a relatively low cost, low risk, non-invasive, highly effective procedure).
(6) Nurses working in a drug and alcohol treatment and rehabilitation facility who scored higher in implicit bias against intravenous drug users experienced more occupational stress, less job satisfaction, and were more likely to express intentions to leave their jobs.	von Hippel et al. (2008)	SC-IAT (attitudes toward intravenous drug users)	Occupational stress, job satisfaction, turnover intentions	Implicit (but not explicit) bias mediated the effect of job stress on turnover intentions.

Table 1 (Continued)

Brief summary (1–2 sentences)	Source/citation	Measure of implicit bias	Behavioral outcome	Research background/elaboration/replications/extensions
(7) Undecided voters' implicit candidate preferences, obtained 1 month prior to the election, significantly predicted their eventual voting decisions.	Arcuri et al. (2008)	IAT (preferences for left- vs. right-wing candidate)	Self-reported voting decisions	The effect was demonstrated in both local and national elections in Italy.
(8) Hazardous drinkers' implicit attitudes toward alcohol predicted heightened appetitive responses in the presence of alcohol and self-reported binge drinking.	Palfai and Ostafin (2003)	IAT (approach vs. avoidance of alcohol)	Reactivity to alcohol cues, binge drinking	Ostafin and Palfai (2006) demonstrated that implicit attitudes predicted binge drinking above and beyond explicit attitudes. Several other replications and extensions exist with regard to alcohol and cocaine use.
(9) Convicted male pedophiles were found to exhibit an implicit association between children and sex, whereas no such effect was observed for other sex offenders (who were not pedophiles).	Gray et al. (2005)	IAT (associations linking children, sex)	Pedophilia (vs. other sex crimes)	Although researchers do not recommend using implicit methods as diagnostic tools, signal detection analysis revealed that the IAT could reliably distinguish pedophiles from non-pedophiles (but it did not perform as well as a commonly used phallometric test).
(10) Adolescents' self-injury implicit associations were correlated with whether or not they had attempted suicide as well as suicidal ideation up to 6 months after their initial assessment, even after adjusting for demographic and psychological risk factors.	Nock and Banaji (2007)	IAT (associations linking self, injury)	Self-reported suicidality	The self-injury IAT enabled investigators to distinguish between a randomly selected individual with no suicidal ideation or attempt history and a randomly selected suicide attempter at a fairly impressive level of accuracy (74–77% correct).

Note: IAT, "Implicit Association Test"; "SC-IAT", "Single Category Implicit Association Test."

## 2.1. Managerial decisions concerning callback interviews

A pair of behavioral economists garnered widespread media attention when they reported the results of an audit study in which they submitted 5000 bogus resumes in response to over 1300 employment advertisements published in Boston and Chicago area newspapers in the following job categories: sales, administrative support, clerical, and customer services (Bertrand & Mullainathan, 2004). The resumes submitted were either relatively high or low in prior work experience and resume quality, and they bore names that were pre-tested to be stereotypical either of European Americans (e.g., Brendan, Todd, Meredith) or African Americans (e.g., Darnell, Rasheed, Tamika). Results of this experiment revealed that job candidates with white names were 50% more likely than equally qualified candidates with black names to receive invitations for job interviews. The magnitude of discrimination was relatively constant across occupation, industry, employer size, and city.

It seems unlikely that the huge racial gap in callbacks demonstrated by Bertrand and Mullainathan (2004) could be explained by the low levels of explicit, self-reported prejudice captured by public opinion surveys cited by Tetlock and Mitchell (2009). However, one cannot be perfectly certain that the discriminatory behavior exhibited by the 1,300 employers in Boston and Chicago was a function of implicit rather than explicit racial biases. Fortunately, two subsequent studies provide more direct evidence that the kinds of race-based hiring biases identified by Bertrand and Mullainathan are linked to implicit prejudice. First, in a follow-up study, Bertrand, Chugh, and Mullainathan (2005) found that scores on an implicit stereotyping task involving race and intelligence were correlated with students' likelihood of selecting resumes with African American names, especially among participants who felt rushed while completing a resume selection task.

Second, Rooth (2007) conducted an illuminating field study using human resource personnel as participants. He examined whether job applicants were contacted for interviews and also administered the IAT to employment recruiters in a Swedish replication and extension of the Bertrand and Mullainathan (2004) study. Using either common Arab or Swedish male names, Rooth (2007) submitted a series of otherwise comparable applications for several different job openings in Stockholm and Gothenburg. The occupations were selected to be highly skilled or unskilled, and they varied in the extent to which they were commonly held by Arabs (e.g., teachers, accountants, restaurant workers, and motor vehicle drivers). From a total of 1552 submitted applications, in 522 cases at least one applicant was contacted by the employer and invited for an interview. When only one applicant was contacted, 217 times this candidate were Swedish (42%) and only 66 times was he or she an Arab (13%); thus, Swedish applicants were three times more likely than Arab applicants to be offered interviews.<sup>10</sup>

Several months later, Rooth (2007) located a subset (26%) of the specific managers responsible for making recruitment decisions for these jobs and paid them to participate in a study “about the recruitment process” in which both implicit and explicit stereotyping measures were administered. Results indicated that managers' scores on an Arab-Swedish IAT—in which the same names used in the job applications were paired with words that were stereotypic of Arabs (e.g., *lazy*, *slow*) vs. white Swedes (*hardworking*, *efficient*)—significantly predicted their likelihood of providing interview opportunities to Arab applicants ( $\beta = -.12$ ,  $p < .05$ ,  $n = 136$ ). A single standard deviation increase in IAT scores was associated with a 12% reduction in the probability of inviting an Arab applicant for an interview. By contrast, explicit measures of stereotyping and prejudice were not predictive of interview decisions.

## 2.2. Hiring preferences, budget cuts, verbal slurs, social exclusion, and physical harm

In encouraging legal and organizational researchers and practitioners to reject the case for implicit bias, Tetlock and Mitchell (2009) dismiss *a priori* any studies in which the research participants are not themselves office managers—as do Wax and Tetlock (2005), Mitchell and Tetlock (2006), Tetlock and Mitchell (2008), and Wax (2008). We believe that dismissing all studies in which non-managers serve as research participants is unwise, unwarranted, and unscientific, given that *no research exists to sustain Tetlock and Mitchell's buoyant assumption that managers would be immune to the effects of implicit bias*. Furthermore, it is unreasonable to demand that all translational research be conducted within a “real” organizational setting, especially given that such a standard would require business

<sup>10</sup> In 239 cases, both Swedish and Arab applicants were contacted.

organizations to countenance potential legal, financial, and public relations risks should employment discrimination be documented in their organization.

In fact, there is a long and distinguished history of translating conclusions from basic social psychological research—including conclusions about human error, motivation, and bias—into managerial and organizational settings (see, *inter alia*, Bazerman & Moore, 2008; Brief, Dietz, Cohen, Pugh, & Vaslow, 2000; Kahneman et al., 1982; Locke, 1986; March & Simon, 1958; Mook, 1983; Pfeffer, Cialdini, Hanna, & Knopoff, 1998; Pfeffer, 1998; Staw, 1991). Given this successful history of applying social psychology to organizational settings, it seems to us that the burden of proof should be on those who would argue that the psychological (and perhaps even physiological) processes of business people are somehow unique. Furthermore, audit studies reveal that racial discrimination continues to exist in the marketplace (e.g., Ayres, 2001; Pager, 2003; Riach & Rich, 2002), and implicit bias could help to explain its continued prevalence. The field studies by Bertrand and Mullainathan (2004) and Rooth (2007) that we have already described—as well as investigations involving police officers, medical doctors, and nurses as research participants (described below)—indicate that professionals are subject to the same kinds of implicit and explicit biases that afflict laypersons.

If it is true that managers at least sometimes behave like other decision-makers, then a set of studies by Rudman and her colleagues warrant serious attention on the part of organizational researchers and practitioners. Rudman and Glick (2001) found that although research participants regarded female managerial applicants who presented themselves as confident, competitive, and ambitious (i.e., agentic) as highly qualified for a leadership role, the participants still discriminated against them (relative to men who were identically described) with respect to hiring decisions because the agentic women were perceived as dislikable and therefore thought to be deficient with respect to social skills. Furthermore, the likelihood of discrimination was increased to the extent that participants possessed implicit gender stereotypes concerning agentic (vs. “communal”) characteristics (see also Jost & Kay, 2005). These findings suggest that the prescription for female “niceness,” which is often internalized at an implicit, unexamined level of awareness, penalizes agentic women in the workplace (see also Eagly & Karau, 2002; Fiske, Bersoff, Borgida, Deaux, & Heilman, 1991; Heilman, Wallen, Fuchs, & Tamkins, 2004; *Price-Waterhouse v. Hopkins*, 1989).

Rudman and Ashmore (2007) conducted two studies that specifically address the criticism raised by Arkes and Tetlock’s (2004) and repeated by Tetlock and Mitchell (2009) that measures of implicit bias have yet to predict behavior that is overtly hostile (i.e., motivated by racial animus). Rudman and Ashmore gauged implicit negative stereotypes using the IAT and found that these were related to the likelihood of performing behaviors that are unambiguously harmful toward minority groups. In the first study, implicit bias scores significantly predicted self-reported racial discrimination, including verbal slurs such as expressing racially or ethnically offensive comments and jokes ( $r = .34, p < .01, n = 64$ ), excluding others from social gatherings and organizations because of their ethnicity ( $r = .30, p < .05, n = 64$ ), and engaging in threat, intimidation, nonverbal hostility (e.g., giving ‘the finger’), and even physically harming members of minority groups and/or their property ( $r = .25, p < .05, n = 64$ ). In the second study, participants were asked for their recommendations about how to spread university budget cuts across several student organizations, including the marching band, drama club, political action groups, and, as the focal dependent variables, Jewish, Asian, and black student organizations. Rudman and Ashmore found that implicit bias scores predicted the likelihood that participants would engage in economic discrimination by disproportionately slashing budgets for the Jewish ( $r = .38, n = 89, p < .05$ ), Japanese ( $r = .30, n = 89, p < .05$ ), and black ( $r = .18, n = 126, p < .05$ ) student associations. In both studies, the IAT again possessed strong incremental validity, predicting harmful and discriminatory behavior even after adjusting for explicit racial and ethnic attitudes.

### 2.3. *Police officers’ decisions to shoot*

As in their earlier critiques, Tetlock and Mitchell (2009) suggest that professional training and accountability constraints will somehow vanquish the behavioral expression of implicit bias. We regard this as an unrealistically optimistic position, especially given the complexity of research findings demonstrating that diversity training can backfire (Plant & Devine, 2001; Roberson, Kulik, & Pepper, 2001, 2003; see also Macrae et al., 1994) and that accountability often has no salutary effect on decision-making biases and may even make some biases worse (Lerner & Tetlock, 1999). We will return later in the chapter to the question of whether organizational accountability is a likely panacea against the effects of implicit bias. For now, we simply note that police officers are among the most highly trained and publicly accountable members of the workforce, especially when it comes to their treatment of racial

minorities, and yet several studies show that, they, too, harbor implicit biases that can sometimes affect their judgment and behavior, as we describe next.

It would obviously be unethical for researchers to create or even to surreptitiously monitor circumstances in which racially biased shooting of innocent victims is a real possibility. Social psychologists have therefore devised various computer simulations to measure “shooter bias” among police officers and civilians (e.g., Correll, Park, Judd, & Wittenbrink, 2002; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001; Plant & Peruche, 2005). In one such study, Correll et al. (2007) found that police officers and civilians alike exhibited robust “shooter bias” in terms of their response speed. That is, both groups of participants were (a) quicker to “shoot” armed black suspects than armed white suspects, and (b) slower to refrain from shooting unarmed black than white suspects. Police officers’ expertise did, however, make a difference in other ways; they outperformed civilians in terms of overall speed and accuracy, and they used a less biased subjective criterion to decide whether to shoot a black vs. white person compared to civilians.

Plant and Peruche (2005) observed in a study of 48 patrol officers in Florida that on early trials of a video simulation the officers were significantly more likely to shoot an unarmed suspect when he was black than when he was white,  $t(47) = -3.17, p < .002$ . However, with practice the officers were able to eliminate their racial bias on later trials. Eberhardt, Goff, Purdie, and Davies (2004, Studies 4 and 5) provide additional evidence that police officers process visual information concerning black and white faces of potential criminals in much the same (biased) way that students and other research participants do. Specifically, they found that the officers’ attention was automatically drawn to a black face more than to a white face after the officers had been subliminally exposed to crime-related words (e.g., *violent* or *arrest*). A follow-up study revealed that cops were also more likely to judge black than white individuals as looking “criminal,” and this was especially true for individual targets who were rated as more race stereotypic in their appearance. These results and others suggest that while training and accountability may mitigate the expression of implicit bias, they do not by any means eradicate all traces of bias, as Tetlock and Mitchell (2009) suppose. In terms of the predictive validity of implicit measures of stereotypical associations, Glaser and Knowles (2008) found that people who implicitly associated blacks with weapons tended to show a stronger “shooter bias.”

#### 2.4. *Medical decisions concerning patient treatment and job turnover*

Business managers and police officers are not the only groups of experts whose implicit biases have been found to affect their professional judgment in consequential ways. For years, medical researchers have been trying to determine why white patients with symptoms of myocardial infarction are nearly twice as likely as blacks with similar symptoms to receive thrombolytic therapy, which is a relatively low cost, low risk, highly effective, non-invasive procedure used to prevent heart attacks. To address this practical question, Green et al. (2007) conducted a landmark study involving 220 medical doctors completing their residencies at several hospitals in Atlanta and Boston.<sup>11</sup> Specifically, the physicians completed measures of implicit and explicit racial bias and were randomly assigned to make a hypothetical diagnosis and expert recommendation about a 50-year-old male patient who happened to be either white or black. Although the diagnoses were hypothetical and the patients’ cases were presented in written form rather than face-to-face, it stands to reason that the doctors would be highly motivated to be accurate in their recommendations and that in rendering their clinical judgments they would draw on sources of information that would be similar to what they would use in treating their own patients.

The results of the Green et al. (2007) study should be of interest to anyone working in the field of healthcare and to citizens (i.e., potential patients) in general. On average, physicians were more likely to recommend thrombolysis for whites (even when they did not believe that the patient was suffering from coronary artery disease), and they were less likely to recommend it for blacks when they did believe the patient was suffering from coronary artery disease. Although they reported no explicit preference for white or black patients or differences in perceived cooperativeness, the white physicians on average exhibited significant pro-white/anti-black biases at the implicit level. Most importantly, with regard to the predictive validity of implicit measures, the degree of pro-white implicit bias reflected

<sup>11</sup> Actually, 287 doctors completed all of the study materials, but 67 expressed some awareness of the researchers’ hypothesis, so they were excluded from the main analyses. Interestingly, it appears that those physicians who were aware that the patient’s race could influence their judgment attempted to overcome their implicit bias in making recommendations, suggesting that the study may have served an “unconsciousness-raising” function for some participants.

in physicians' scores on the race IAT was negatively associated with recommending thrombolysis for black patients ( $\beta = -.19, p < .05$ ) and positively associated with recommending thrombolysis for white patients ( $\beta = .17, p < .01$ ). The statistical interaction between implicit bias and patient race on treatment decisions remained significant even after adjusting for physicians' demographic characteristics (race, sex, and SES), degree of explicit racial bias, and confidence in the effectiveness of thrombolysis. This study demonstrates that implicit racial bias predicts the withholding of valuable medical treatment for some patients, suggesting that one's degree of implicit bias can have life-or-death consequences for others.

In another study of the implicit attitudes of medical personnel, von Hippel, Brener, & von Hippel (2008) focused on nurses working in a drug and alcohol treatment and rehabilitation facility in Sydney, Australia. Results indicated that nurses who scored higher in implicit bias against intravenous drug users experienced more occupational stress ( $r = .40, n = 44, p < .01$ ) and less job satisfaction ( $r = -.33, n = 44, p < .05$ ), and they were also more likely to express intentions to leave their jobs ( $r = .51, n = 44, p < .01$ ), compared to nurses who were lower in implicit bias. These effects were robust against statistical outliers and held after adjusting for explicit attitudes concerning intravenous drug users. Importantly, von Hippel et al. found that implicit (but not explicit) bias mediated the effect of stress on turnover intentions. Thus, implicit attitudes predict not only how medical professionals treat their patients but also how treating certain kinds of patients affects the work experiences and daily lives of medical professionals.

### 2.5. Undecided voters: for whom will they (eventually) vote?

Although its specific relevance to business managers may be limited, another research program focusing on “real world” decision-makers has set its sights on voters casting ballots in local and national elections in Italy (Arcuri, Castelli, Galdi, Zogmaister, & Amadori, 2008). As Arcuri et al. (2008) point out, undecided voters comprise a sizable portion of the electorate and thus frequently play a decisive role in elections, but their behavior is characterized by a high degree of uncertainty; many of the “undecided” reach a decision only a few days before election day, or even while they are in the polling booth. In two studies, the researchers investigated whether these voters' eventual decisions could be predicted in advance by their implicit attitudes despite the fact that they profess no candidate preference at the explicit level. In Study 1, in advance of the 2001 national election, voters in Milan reported their explicit preference for one of two major candidates (Silvio Berlusconi, leader of the right-wing coalition, or Francesco Rutelli, leader of the left-wing coalition) and performed an IAT that paired the two politicians with pleasant and unpleasant attributes. Four weeks later, participants indicated how they had actually voted. Restricting analyses to those 52 participants who voted for one of the two major candidates (including 33 who had already decided and 19 who had not yet decided at the time of the initial session), the authors found that participants' IAT scores were highly predictive of their voting behavior,  $Wald(1) = 12.73, p < .001$ .<sup>12</sup>

In a second study, Arcuri et al. (2008) recruited 58 undecided voters (age range: 22–65 years) a month before local elections were held in northern Italy. After performing the candidate IAT, respondents were given materials to mail in after the election so they could indicate how they had actually voted. For those who voted for either of the two candidates and returned the study materials ( $n = 37$ ), IAT scores reliably predicted eventual voting decisions,  $\chi^2(1) = 6.33, p < .02$ . The authors concluded that some undecided voters may have vigorous implicit preferences even when they are unaware of them, and their implicit preferences exert a significant influence on behavior inside the voting booth. If true, this suggests that assessing implicit attitudes could improve election (and perhaps other types of) forecasts whenever a substantial proportion of voters (or respondents) claim to be undecided.

### 2.6. Binge drinking, pedophilia, and suicide attempts

Several studies indicate that implicit attitudes are valid, useful predictors of self-reported behaviors that are extremely consequential from the standpoint of mental and physical health as well as from the standpoint of others who are affected by those behaviors, even if they are not all directly applicable to managerial decision-making. For instance, Palfai and Ostafin (2003) administered a modified approach/avoidance IAT to measure implicit attitudes

<sup>12</sup> This result is reminiscent of the robust correlation observed between candidate IAT scores and explicit voting preferences during the 2000 U.S. presidential election between George W. Bush and Al Gore for thousands of Web site respondents (Nosek, Smyth, et al., 2007).

concerning alcohol to a sample of 48 hazardous drinkers. They found that IAT scores reflecting a strong alcohol-approach propensity were associated with self-reported frequency of binge drinking ( $r = .34$ ,  $n = 47$ ,  $p < .05$ ). Furthermore, implicit attitudes predicted participants' reactivity to alcohol cues, that is, the extent to which they reported experiencing appetitive responses when a glass of beer was placed before them. These findings were replicated and extended by Ostafin and Palfai (2006), who demonstrated that implicit attitudes concerning alcohol possess incremental validity; in other words, IAT scores predict binge drinking above and beyond explicit, self-reported attitudes concerning alcohol use. Several other replications and extensions can be found in the published research literature on drug and alcohol addiction (e.g., Thush & Wiers, 2007; Thush et al., 2007; Wiers, Van Woerden, Smulders, & De Jong, 2002; Wiers, Houben, & de Kraker, 2007).

An even more startling clinical application of research on implicit attitudes concerns the case of pedophilia. For obvious reasons, pedophiles are strongly motivated to conceal their sexual attraction to children, which means that explicit measures are of relatively little diagnostic and forensic use. Nevertheless, a study of 78 male sexual offenders in a Welsh prison revealed that implicit associations, as measured in terms of IAT scores, were indeed correlated with being indicted for sexual assault against children (Gray, Brown, MacCulloch, Smith, & Snowden, 2005). Researchers found that convicted pedophiles exhibited an implicit association between children and sex, whereas other sex offenders (who were not pedophiles) did not; differences between the two criminal groups were large in terms of conventional measures of effect size (Cohen's  $d = .84$ ). Gray et al. also used signal detection analysis to estimate the proportion of hits to false alarms that would be produced if one were to use the IAT as a diagnostic tool and concluded that, although it did not perform quite as well as a phallometric test for pedophilia, the IAT exhibited good potential for distinguishing pedophiles from non-pedophiles.

There is evidence, too, that implicit attitudes are correlated with suicidal ideation and actual suicide attempts by at-risk adolescents. Nock and Banaji (2007) administered to 89 adolescents a self-injury IAT, which measured the strength of association between the self-concept and the desire to harm oneself. Adolescents with a history of suicide attempts produced significantly higher self-injury IAT scores than did a group of adolescents with no such history. Furthermore, the self-injury IAT enabled investigators to distinguish between a randomly selected individual with no suicidal ideation or attempt history and a randomly selected suicide attempter at a fairly impressive level of accuracy (74–77% correct). Finally, statistical analyses revealed that even after adjusting for demographic and psychological risk factors, the self-injury IAT predicted a significant amount of the variance in suicidal ideation, actual behavioral suicide attempts, and suicidal ideation up to 6 months after the initial assessment.

We challenge critics of the IAT, including Tetlock and Mitchell (2009), to identify any other technique for measuring socially sensitive attitudes or behavioral dispositions (in a relatively nonreactive manner) that has been used successfully to predict such consequential, problematic behavioral outcomes as binge drinking, pedophilia, and suicide attempts. Although these specific outcomes may or may not be considered organizationally relevant (depending on several factors, including the type of organization), they are obviously socially significant. When evidence of this kind exists, it is difficult to understand how critics can so blithely call into question the behavioral significance (and therefore the practical relevance) of implicit associations and evaluations.

### 3. Other problems with the Tetlock–Mitchell critique

Now that we have addressed Tetlock and Mitchell's (2009) chief complaint, namely that measures of implicit bias fail to predict organizationally relevant behavioral outcomes, we turn our attention to several other details of their critique. The remaining objections can be classified as primarily ideological (or, in some cases, perhaps philosophical) or as primarily methodological, and we consider each type in turn. Some, we think, are specious while others deserve to be taken seriously, but they do not—even when taken in conjunction—justify the kind of blanket skepticism that Tetlock and Mitchell advocate. Rather, the body of evidence undermines Tetlock and Mitchell's case for complacency (echoed by Wax, 2008) concerning the societal and organizational relevance of implicit bias.

#### 3.1. A refutation of philosophical and ideological objections

In challenging the notion that implicit bias is a genuine problem that requires thoughtful analysis and perhaps even consideration of redress, Tetlock and Mitchell (2009) follow a campaign-like “kitchen sink” approach, throwing together a wide range of largely unrelated complaints and criticisms that have been leveled against research on implicit

bias, especially research using the IAT. Some of these criticisms may sound like scientific objections, but they are ideological in character. Specifically, Tetlock and Mitchell envision a scientific hell (see also Tetlock, 1994) in which “market purists” (conservatives) and “statist interventionists” (liberals) are “trapped for eternity in a theater-of-the-absurd dialogue of the deaf” (p. 21). Although they do acknowledge that market purists are capable of tuning out, it is clear—especially to readers of Mitchell and Tetlock (2006), Tetlock and Mitchell (2008), and Wax and Tetlock (2005)—that they regard statist interventionists as especially hard of hearing on the subject of implicit bias. In fact, they come dangerously close to advocating a scientific double standard that would privilege anti-egalitarian sympathies, writing that: “we should be all the wari-er when the claims dovetail conveniently with the egalitarian sympathies of the research community” (p. 11).

Tetlock and Mitchell (2009) argue that statist interventionists refuse to accept the evidence from public opinion surveys that prejudice is a thing of the past, that their ideological commitments lead them to ignore and even to distort other scientific facts, and that their goal is some kind of witch-hunt to round up and punish all the implicit racists. They also argue, as they have in several prior publications (e.g., Arkes & Tetlock, 2004; Wax & Tetlock, 2005), that if black liberals (such as Jesse Jackson) were to exhibit anti-black/pro-white biases on implicit measures, such evidence would mean that implicit measures must be measuring something other than prejudice. We respond to each of these claims in turn, pointing out that the last argument in particular is fallacious on several counts and ought to be retired once and for all.

### 3.1.1. *Have you heard the news?*

There are few methodological innovations in social psychology that have inspired Op-Ed pieces in the *Wall Street Journal*, but the Implicit Association Test has. In 2005, Tetlock teamed up with Amy Wax, the outspoken conservative law professor and skeptic about the significance of social scientific research for understanding or improving race and gender relations (e.g., see Wax, 2008). Wax and Tetlock began by suggesting that researchers of implicit bias have ignored the great strides made against prejudice in society:

It was once easy to spot a racial bigot: The casual use of the n-word, the sweeping hostility, and the rigid unwillingness to abandon vulgar stereotypes left little doubt that a person harbored prejudice toward blacks as a group. But 50 years of survey research has shown a sharp decline in overt racial prejudice. Instead of being a cause for celebration, however, this trend has set off an ever more strident insistence in academia that whites are pervasively biased.

Tetlock and Mitchell (2009) similarly charge researchers of implicit bias with underestimating the “sincerity of the tolerant attitudes that many people in early 21st America claim to possess (creating the implication that managers are far more biased than they realize or are willing to admit)” as well as the “sincerity of organizational efforts to check prejudice” (p. 4).

Conscientious readers of the research literature on implicit social cognition and automatic behavior should know that—according to the legions of social and cognitive psychologists studying the role of unconscious mental processes—sincerity (and good intention) has absolutely nothing to do with it. The fact is that many people are sincere in holding egalitarian ideals and yet harbor implicit biases (e.g., Devine, 1989). Nevertheless, one can raise important questions about the *effectiveness* of individual and organizational attempts to regulate stereotyping and prejudice and, in our view, it would be irresponsible *not* to raise such questions, given the state of psychological evidence.

Furthermore, members of disadvantaged groups frequently see less “cause for celebration,” as Wax and Tetlock (2005) put it, with respect to the progress that has been made concerning racism and sexism in society, in comparison with members of advantaged groups. Although most blacks tend to agree that the problem of racial prejudice has improved over the past several decades, they are far less sanguine about the “end of racism” than are whites (e.g., Schuman, Steeh, Bobo, & Krysan, 1997). For example, a 2004 survey revealed that a strong majority of blacks (58%) believe that discrimination is still the main obstacle to racial equality, compared with just 30% of whites. Other major racial differences in public opinion concerning the existence of racial discrimination in education, housing, the job market, and in treatment by the police are summarized in Table 2. These data reveal that African Americans are far more likely than European Americans to believe that discrimination remains an ongoing problem in U.S. society (see also Cohen & Agiesta, 2008; Tyler, Boeckmann, Smith, & Huo, 1997). For example, black respondents are twice as likely to believe that police officers treat blacks unfairly and they are almost five times as likely to believe that blacks are treated unfairly at work, compared to white respondents. Given that members of disadvantaged groups hold much



Table 2  
Racial gap with respect to public opinion concerning the persistence of racial discrimination in the U.S.

Compared with whites, blacks. . .	Percentage of whites saying “yes”	Percentage of blacks saying “yes”
(1) Have an equal chance at education	86%	52%
(2) Have an equal chance at housing	87%	56%
(3) Have an equal chance at jobs	79%	42%
(4) Are treated unfairly at work	11%	53%
(5) Are treated unfairly by the police	35%	70%

*Note:* These data were taken from the 2004 Gallup Poll (“Racial Attitudes in America”) and were downloaded June 27, 2008, from <http://www.igpa.uiuc.edu/programs/racialAttitudes/>. They summarize updated trends since the publication of the revised edition of Schuman et al. (1997). Questions were worded as follows: (1) “In general, do you think that black children have as good a chance as white children in your community to get a good education, or don’t you think they have as good a chance?” (2) “Again, in general, do you think that blacks have as good a chance as white people in your community to get any housing they can afford, or don’t you think they have as good a chance?” (3) “In general, do you think blacks have as good a chance as white people in your community to get any kind of job for which they are qualified, or don’t you think they have as good a chance?” (4) “Just your impressions, are blacks in your community treated less fairly than whites in the following situations? How about: on the job or at work?” (5) “How about by police, during traffic incidents?” Response options were either “yes” or “no.”

less rosy views of the state of race relations than do members of advantaged groups, Tetlock and Mitchell’s injunction to take whites’ views of the degree of racism in society (and in themselves) at “face value” is tantamount to ignoring or rejecting blacks’ own perceptions concerning the persistence of discrimination.<sup>13</sup>

It would be foolhardy to measure intelligence (and many other psychological constructs) by explicit self-report, as if asking people whether they *think* they are smart provides reliable evidence about their actual degree of intelligence (e.g., Banaji, 2001). Why, then, should we rely on whites’ reports of their own prejudice levels, as Tetlock and Mitchell propose? Even if people were motivated to tell the truth (and were unaffected by social desirability concerns), explicit measures can only reveal what people *believe* about their attitudes. This may be valuable information in some cases, but it is certainly not the whole story, and still less grounds for “celebrating” the “end of racism.” An understanding of implicit processes helps to explain the significant gap between the (sincerely) egalitarian claims of whites and the impression among most blacks that racial inequality persists anyway.

### 3.1.2. Facts have nothing to do with it

The articles by Tetlock and Mitchell (2009) and Wax and Tetlock (2005) contain a criticism of implicit bias researchers that is much sterner than the charge of unwarranted pessimism. It is that “statist interventionist” (i.e., “liberal”) ideology has led researchers to distort the scientific facts, exaggerate the case for the existence of implicit bias, and to insinuate themselves in policy debates where they do not belong. Wax and Tetlock conclude their editorial with a dressing down:

But facts have nothing to do with it. What began as science has morphed into unassailable faith. However we think, feel or act, and however more apparent progress has been made, there is no hope for us. We are all racists at heart.

Tetlock and Mitchell (2009) are a bit more polite than this, and their analysis of ideological disputes concerning the existence of implicit bias is dressed up in some pretty fancy (and highly readable) philosophy of science, but the take-home message is the same. They assert that there is a “mismatch between empirical accomplishments and policy prescriptions” (p. 5), that an “expansive conception of prejudice is presumptuous, a usurpation of the political by the psychological” (p. 23), and that “proclamations” by implicit bias researchers could “damage the long-term credibility of social science” (p. 6). In other words, researchers of implicit bias are not only wrong but also irresponsible.

Our philosophy of science is not nearly as skeptical as that of Tetlock and Mitchell (2009), who suggest that the scientific status quo leaves both sides in the debate over implicit bias “free to inflate or deflate its estimates of prejudice

<sup>13</sup> Tetlock and Mitchell (2009) may also be overstating whites’ optimism with regard to race relations. A *Washington Post*-ABC News poll found that “nearly half of all Americans say race relations in the country are in bad shape and three in 10 acknowledge feelings of racial prejudice” (Cohen and Agiesta, 2008, p. A01).

in response to the same facts” (p. 5). Furthermore, we dispute their assertions that there is a “paucity of probative data” or an “absence of shared standards of evidence” within the professional community of social scientists. We have already summarized several studies that have indeed produced probative data according to shared standards of evidence, and we trust conscientious readers to decide the empirical case on its merits rather than their own biases (or ideological proclivities). There is a scientific consensus; it is just that Tetlock and Mitchell have decided to abstain.<sup>14</sup>

Most emphatically, we reject Wax and Tetlock’s (2005) cynical claim that “facts have nothing to do with it.” On the contrary, the facts that we review in this article provide an overwhelming case for the existence of implicit bias, and this case is not seriously questioned by social psychological experts in general. These facts also provide enough evidence for reasonable third-parties, perhaps even some openminded “market purists,” to conclude that—to the extent that managers share at least some of the psychological characteristics of students, nurses, doctors, police officers, and most of the rest of the population—the likelihood that implicit bias contributes to workplace discrimination is, if not a perfect certainty, close enough.

### 3.1.3. *Caricaturing the goals and assumptions of implicit bias research*

There are a number of ways in which Tetlock and Mitchell (2009) caricature or otherwise misrepresent the arguments made by researchers of implicit bias. For example, they falsely ascribe to such researchers the belief that egalitarian behavior at work is *impossible*, including the assumption that “prejudice so suffuses American life that it is impossible for managers to make color-blind decisions in any discretionary phase of personnel decision-making in which protected-category status is known or can be inferred” (p. 7). But, as we have already mentioned with respect to research programs addressing implicit and explicit motivation to control prejudice, one of the central goals of research on implicit intergroup attitudes is to identify both personal and situational factors that make egalitarianism more likely and its opposite, intergroup bias, less likely (e.g., Amodio et al., 2003, 2004, 2006, 2008; Blair, 2002; Dasgupta & Rivera, 2006; Devine et al., 2002; Dunton & Fazio, 1997; Fazio et al., 1995; Glaser & Knowles, 2008; Hofmann, Gschwendner, Castelli, & Schmitt, 2008; Lowery et al., 2001; Moskowitz et al., 1999; Park et al., 2008; Plant & Devine, 1998; Sinclair et al., 2005). Tetlock and Mitchell (2009) also claim that researchers of implicit bias assume that “unconscious bias is so pervasive that only de facto quota systems can keep it at bay” (p. 30), but many other strategies for overcoming implicit bias have been suggested in the research literature (and, to our knowledge, quotas have not). These include cultivating egalitarian motives (Dasgupta & Rivera, 2006; Moskowitz et al., 1999), exposing people to favorable, counterstereotypical exemplars (Dasgupta & Asgari, 2004; Dasgupta & Greenwald, 2001; Lowery et al., 2001), providing opportunities for emotional reconditioning (Rudman, Ashmore, & Gary, 2001), increasing vigilance about one’s subtle behavior during interactions with disadvantaged others (Dasgupta & Rivera, 2006), and educating people about their implicit biases (i.e., “unconsciousness raising”; Green et al., 2007; Pronin, 2007).

Regrettably, it is becoming increasingly common among certain critics of the IAT—in the absence of a broadly convincing methodological critique—to impugn the motives of the technique’s developers, insinuating without evidence that the goal of implicit bias research is to generate some sort of witch-hunt against implicit racists (or sexists or whomever). For example, Tetlock and Mitchell (2009, p. 25) misleadingly suggest that the guiding question of implicit bias research is: “When do we have warrant to call people prejudiced?”<sup>15</sup> Their language of “false accusations” of racism and people who either “pass” or “fail” the IAT is inflammatory and inaccurate as a description

<sup>14</sup> To our surprise, Tetlock and Mitchell (2009) also seem to reject the consensus among social psychologists concerning several other disciplinary staples, including the fundamental attribution error, belief in a just world, and stereotype threat (see p. 7). They also appear (p. 10) to depart from received wisdom since Nisbett and Wilson (1977) that private introspection is insufficient as a basis for inferring the causes of one’s own behavior. Finally, Tetlock and Mitchell mock the notion that nonverbal behavior could be (in some cases) indicative of prejudicial attitudes (p. 4), although this idea has a long and venerable tradition within social psychology that can be traced at least as far back as Bogardus’ (1925) pioneering work on prejudice and social distance and to Word, Zanna, and Cooper’s (1974) classic experimental demonstrations of how bias can affect nonverbal behavior in interviewing situations (see also Amodio and Devine, 2006; Dasgupta and Rivera, 2006; Dovidio et al., 1997, 2002; Fazio et al., 1995; McConnell & Leibold, 2001). Given the existing evidence base, we regard Tetlock and Mitchell’s generic dismissal of nonverbal research on prejudice and bias to be unscientific and again out of step with the social psychological community. They are also factually incorrect when they accuse researchers of implicit bias of treating “eye-blinking” as “evidence of prejudice” (p. 4).

<sup>15</sup> It is important to point out that Tetlock, Mitchell, and their co-authors have not provided any direct evidence (textual or otherwise) of IAT researchers’ alleged ideological bias. As a result, their ideological critique amounts to a series of *ad hominem* attacks. Innuendo may be commonplace in the opinion pages of the print news media and especially the Blogosphere, but it is disappointing to see the intellectual standards of the legal and scientific periodicals in which their articles have been published flouted in this way.

of the goals and motives of a team of conscientious scientists who have responded in good faith (not to mention good humor) to increasingly strident attacks by conducting extensive statistical analyses (e.g., Greenwald et al., 2009), proposing methodological and scoring improvements (Greenwald, Nosek, et al., 2003; Greenwald, Oakes, et al., 2003), and repeatedly cautioning *against* the use of the IAT as a selection instrument to “diagnose prejudice” on an individual basis (e.g., Nosek, Greenwald, et al., 2007; this opinion was offered under oath by Banaji in testimony provided in *New Hampshire v. Michael Addison*, 2009). Indeed, researchers have first and foremost regarded implicit bias as an unfortunate social psychological consequence of hierarchical social systems and only secondarily as a problem of the individual (e.g., Hardin & Banaji, *in press*; Jost & Banaji, 1994).

We regard Tetlock and Mitchell’s accusations of ideological bias, therefore, as unjust and note that they are belied by the educational mission of the Project Implicit website (<https://implicit.harvard.edu/implicit/>) and the extensive community outreach undertaken for several years by Mahzarin Banaji, Tony Greenwald, and Brian Nosek, among others. These researchers and others in the field have—far from accusing others of bigotry or hypocrisy—shared their own personal experiences (and even their own IAT scores) in an effort to elevate the level of discourse concerning prejudice in society so that the debate is not about finger-pointing but rather about facilitating opportunities for individuals to consider for themselves the possibility that they may have internalized the residue of longstanding histories of racial and other forms of discrimination in the culture-at-large. As educators who have used the IAT to spur dialogue in our own classrooms, we find it to be unsurpassed as a pedagogical tool for critical self-reflection with respect to intergroup attitudes. Students who begin with the impression that racism, sexism, and other forms of prejudice no longer pose a threat to equal opportunity ideals often discover new ways of appreciating the situations faced by members of disadvantaged minority groups. Of course, some people respond defensively in response to personalized IAT feedback, but this, too, may be a necessary concomitant of introspection, at least for a subset of individuals.

In any case, students are not alone in their ability to be enlightened by exposure to research on implicit bias. The physicians who participated in the study by Green et al. (2007) also praised the educational value of the IAT. They were asked whether they agreed with the statement, “Subconscious biases about patients based on their race may affect the way I make decisions about their care without my realizing it” either before or after taking the IAT. When they answered before taking the IAT, 61% of the doctors agreed; when they responded after taking the IAT, 72% agreed (a significant increase). Moreover, three-quarters of the participants (75%) agreed that taking the IAT is a worthwhile experience for physicians, and that learning more about unconscious biases could improve the quality of their patient care.

In summary, then, the IAT seems to be a remarkably effective tool for both illuminating and educating people about the subtle ways in which inequality in the social system is perpetuated by individuals, often outside of their own awareness. We think that Arkes and Tetlock (2004, p. 275) got it exactly right when they noted that:

If we think of racial prejudice as the primal blemish on America’s collective reputation as a just society, and if we agree with de Tocqueville (1835/2001) that Americans have a uniquely ‘lively faith in the perfectability of man’ (p. 359), the IAT can be viewed as a quintessentially grass-roots project to use the best available science to expunge racist sentiments not only from the consciousness of Americans but from their unconscious as well.

If only they had left it at that.

### 3.1.4. *The parable of the “Two Jesses” (once more with feeling)*

Tetlock and Mitchell (2009) reprise the eponymous example from the Arkes and Tetlock (2004) critique, which was entitled “Attributions of Implicit Prejudice, or ‘Would Jesse Jackson ‘Fail’ the Implicit Association Test?’” This “parable,” which centers on the authors’ speculations about how (black liberal) Jesse Jackson and (white conservative) Jesse Helms would “perform” on a race-based IAT, was also featured prominently in the Wax and Tetlock (2005) Op-Ed piece that chided researchers of implicit bias for allegedly concluding that, “We are all racists at heart.” Tetlock and Mitchell (2009) write that:

One Jesse (Jackson) is a statist interventionist who declares discrimination to be an ongoing, not just historical problem—many Whites still resent African-Americans. The other Jesse (Helms) is a market purist who believes that the big causes of racial inequality in America are now internal to the African-American community, especially the erosion of responsibility in inner cities. . . . Some day, someone may offer a compelling reason to expect these two individuals to exhibit different reaction times on the IAT. But no one has yet.

Table 3

Race-based IAT means as a function of respondent's race and political orientation.

	Extremely liberal	Somewhat liberal	Slightly liberal	Centrist/moderate	Slightly conservative	Somewhat conservative	Extremely conservative
European Americans							
Mean	.37	.47	.52	.54	<b>.57</b>	<b>.60</b>	<b>.61</b>
SD	.50	.49	.49	.47	.47	.47	.48
N	1388	2291	1277	2963	1082	1226	417
African Americans							
Mean	<b>-.02</b>	<b>.09</b>	<b>.07</b>	.01	.04	.00	-.06
SD	.55	.57	.58	.54	.55	.59	.55
N	181	350	197	490	120	89	37

Note: These data are adapted from Jost et al. (2004). The means that are most relevant for evaluating the claims made by Arkes and Tetlock (2004), Tetlock and Mitchell (2009) and Wax and Tetlock (2005) that black liberals (such as Jesse Jackson) are likely to score similarly to white conservatives (such as Jesse Helms) are boxed and boldfaced.

As a matter of fact, Jost et al. (2004, pp. 902–904) presented data from over 700 African American liberals and over 2500 European American conservatives, and their results certainly do provide a “compelling” statistical reason to expect that Jesse Jackson and Jesse Helms would differ on a race-based IAT. The data are summarized in Table 3. Whereas white conservatives (on average) exhibited a strong pro-white/anti-black bias, black liberals (on average) hovered around the zero-point, suggesting little or no implicit bias overall.<sup>16</sup>

The stark difference that can be seen in Table 3 between the IAT responses of “extremely conservative” white respondents and “extremely liberal” black respondents ( $d = 1.31$ , a very large effect size for the difference) altogether refutes Tetlock and Mitchell’s case, which was, after all, based on little more than their own creative embellishment of a single remark ascribed to Jesse Jackson.<sup>17</sup> Arkes and Tetlock (2004) doubted that reaction-time measures could ever reliably “differentiate people [such as the Two Jesses] who share a considerable knowledge base but who differ only in their causal attributions for between-group inequality” (p. 264), but the research by Jost et al. (2004) shows quite clearly that such ideological differences are indeed picked up by the IAT.<sup>18</sup> That Tetlock, Mitchell, and their collaborators have repeatedly ignored the data that are most relevant to their speculations about how white conservatives and black liberals might respond on a race-based IAT raises serious concerns about the comprehensiveness of the literature reviews on which their critiques are based. One of the purposes of our “Executive Summary” has been to highlight studies that are overlooked by skeptics in their haste to dismiss the case for implicit bias.

<sup>16</sup> Blanton and Jaccard (2006) assert that the zero-point in IAT research does not necessarily reflect “no bias” (i.e., egalitarianism), but they provide no pertinent data to support this assertion. Implicit bias researchers have responded with evidence that bears on this issue (Greenwald, Nosek, & Sriram, 2006; Greenwald, Rudman, Nosek, & Zayas, 2006). In any case, it seems clear enough that scores around zero reflect less pro-white/anti-black bias than scores ranging from .57 to .61 (see Table 3).

<sup>17</sup> The remark, which served as the epigram for Arkes and Tetlock (2004), was this: “There is nothing more painful to me at this stage in my life than to walk down the street and hear footsteps and start thinking about robbery. Then look around and see somebody White and feel relieved.”

<sup>18</sup> There is a sense, then, in which the data reported by Jost et al. (2004) provide another piece of evidence in support of the validity of the IAT as a measure of racial attitudes, insofar as political orientation (liberalism–conservatism) shows the same pattern of association with implicit bias as with explicit bias. That is, increasing conservatism in whites is associated with significant linear increases in pro-white/anti-black bias on both implicit (Jost et al., 2004) and explicit measures of intergroup attitudes (*inter alia*, Sears et al., 1997; Sidanius et al., 1996).

But we think that there is an even bigger problem with the “Two Jesse’s” argument, and it suggests a truly fundamental misunderstanding concerning the nature of prejudice. Tetlock and Mitchell (2009) assume—as did Arkes and Tetlock (2004, p. 262)—that people cannot be *prejudiced* against members of their *own* group. But this assumption ignores over 60 years of scholarship in social psychology, including the work of Lewin (1941), who analyzed the problem of “Jewish self-hatred,” and Allport (1954), who noted that “so heavy is the prevailing cultural pressure that members of minority groups [such as African Americans] sometimes look at themselves through the same lens as other groups” (p. 198; see also Jost & Hamilton, 2005). This phenomenon has also been studied rather extensively in recent years by system justification researchers, who have found that it is quite common for members of disadvantaged groups to internalize social and cultural biases against their own kind and, consequently, to devalue their own group relative to more advantaged outgroups on both implicit and explicit measures of intergroup attitudes (e.g., Ashburn-Nardo, Knowles, & Monteith, 2003; Jost, Pelham, & Carvallo, 2002; Jost et al., 2004; Rudman, Feinberg, & Fairchild, 2002; Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). Thus, even if Arkes and Tetlock (2004) had been correct that a given black respondent (such as Jesse Jackson) or black respondents in general were prone to display evidence of a pro-white/anti-black bias on implicit measures, the conclusion simply does not follow that the measures must be tapping something other than stereotyping and prejudice.

### 3.2. A refutation of methodological objections

In this section, we address three remaining methodological objections leveled by Tetlock and Mitchell (2009) against work on implicit bias. All of these objections have appeared in one form or another in previous publications (e.g., see Arkes & Tetlock, 2004; Blanton & Jaccard, 2006; Karpinski & Hilton, 2001; Mitchell & Tetlock, 2006; Tetlock & Arkes, 2004; Tetlock & Mitchell, 2008), so it would be desirable to obtain closure on these issues, in order to save both sides in this debate from being “trapped for eternity in a theater-of-the-absurd dialogue of the deaf,” as Tetlock and Mitchell (2009, p. 21) so dramatically put it. We start by revisiting the question of whether implicit biases are likely attributable to “benign” causes such as unfamiliarity, cultural awareness, sympathy, fear of being labeled a bigot, or cognitive dexterity. Afterward, we weigh in on the question of whether organizational accountability is likely to stop implicit bias from influencing judgment and decision-making at work, as Tetlock and Mitchell suggest.

#### 3.2.1. Are implicit biases due to “benign” causes?

Echoing earlier critiques (e.g., Arkes & Tetlock, 2004), Tetlock and Mitchell (2009) assert (still without any evidence of their own) that many factors other than automatic preferences are capable of influencing IAT scores. We do not doubt that this is true, at least to some extent, but the IAT is hardly unique in this respect; no measure in the behavioral sciences is “pure.” Tetlock and Mitchell again propose (but do not investigate) a number of alternative influences, including one’s degree of familiarity or sympathy with members of target groups, one’s awareness of cultural stereotypes, one’s level of cognitive dexterity, and one’s fear of being labeled a bigot. Our bottom-line response is fairly simple. If IAT scores were measured *nothing* more than familiarity or sympathy (or any of the other artifacts proposed by critics), then there is no way that such scores would predict discriminatory attitudes and behaviors in the manner and to the extent that they do (see also Dasgupta, 2004; Greenwald et al., 2009; Hardin & Banaji, *in press*; Hofmann et al., 2005; Jost et al., 2004; Nosek, 2005; Nosek, Smyth, et al., 2007). Nevertheless, many researchers of implicit bias have gone to the trouble of empirically investigating the role of variables such as familiarity, cultural awareness of stereotyping and prejudice, and so on. Without going into unnecessary levels of detail, we summarize the highlights of these research programs in hopes of undoing the unfortunate impression left by Tetlock and Mitchell that implicit bias researchers have failed to consider obvious artifactual interpretations of their effects.

#### 3.2.2. Familiarity

According to the “mere exposure effect,” people tend to prefer familiar over unfamiliar stimuli (Zajonc, 1968). It follows that when people are more or less familiar with specific target categories and/or exemplars (e.g., black vs. white names or faces), individual differences in familiarity could conceivably account for differences in implicit preferences. However, studies in which experimenters have systematically varied the familiarity of the stimuli or adjusted statistically for observed differences in their familiarity, have ruled out this explanation of implicit bias in general (Amodio et al., 2003) and IAT effects in particular (e.g., Dasgupta, McGhee, Greenwald, & Banaji, 2000;

Dasgupta, Greenwald, & Banaji, 2003; Ottaway, Hayden, & Oakes, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999). The fact that Tetlock and Mitchell (2009) continue to suggest that implicit attitudes could merely reflect one's degree of familiarity with stimulus materials without delving into the most relevant research literature is questionable, to say the least.

For example, researchers from two independent laboratories matched black, Hispanic, and white names both in terms of objective frequency counts and subjective familiarity ratings and found that even under these circumstances whites showed strong implicit pro-white biases (Dasgupta et al., 2000; Ottaway et al., 2001). Rudman et al. (1999) investigated attitudes toward American vs. Russian leaders under four different conditions. In one, the IAT contrasted familiar American leaders (e.g., Lincoln, Kennedy) with unfamiliar Russian leaders (e.g., Suslov, Mikoyan). In another condition, the IAT contrasted unfamiliar American leaders (e.g., Fillmore, Pierce) with familiar Russian leaders (Khrushchev, Lenin). In the two other conditions the stimuli were matched in terms of familiarity (i.e., familiar American and Russian leaders; unfamiliar American and Russian leaders). Contrary to what one would expect if implicit biases were driven by familiarity, the magnitude of the pro-American IAT bias (for U.S. participants) was virtually identical in all four conditions.

### 3.2.3. Cultural awareness

Arkes and Tetlock (2004) argue that implicit bias measures may tap “shared cultural stereotypes rather than personal animus,” without apparently realizing that most implicit bias researchers have been arguing (for nearly 30 years) that (a) it is possible for people to be biased without feeling personal animus (or wanting to be biased in any way), and (b) awareness of shared cultural stereotypes (and objective inequalities among groups in the social system) are indeed major sources of implicit bias (e.g., Banaji, 2001; Banaji et al., 1993; Devine, 1989; Eagly & Karau, 2002; Fiske, 1998; Gross & Hardin, 2007; Haddock, Zanna, & Esses, 1993; Jackman, 1994; Jost & Banaji, 1994; Jost & Hamilton, 2005; Katz & Braly, 1935). Thus, the question of whether implicit attitudes reflect unadulterated “personal,” idiosyncratic preferences (see also Karpinski & Hilton, 2001, Olson & Fazio, 2004) *rather than* the cultural milieu (i.e., “extrapersonal associations”) is ill-posed. As Allport (1954) argued with respect to explicit prejudice, an individual's attitudes almost certainly have both a personal and a cultural basis.<sup>19</sup> Because there is considerable variability in implicit attitudes across individuals within the same culture—and such variability is meaningfully related to variability in discriminatory conduct—it appears that implicit bias is not reducible to “purely” cultural factors.<sup>20</sup>

Nevertheless, Nosek and Hansen (2008) have conducted the most comprehensive test of the “extrapersonal associations” hypothesis to date, administering 95 different attitude IATs to thousands of Project Implicit Website respondents. They found that IAT effects were, overall, not influenced by participants' beliefs about how the culture valued each attitude object (median  $r = .01$ , average sample size = 1053); even the largest correlation was relatively weak ( $r = .15$ ). By contrast, individuals' IAT scores were significantly and positively correlated with their explicit attitudes for all 95 attitude objects, with a median  $r = .36$  (maximum  $r = .70$ ). The researchers conclude, appropriately enough, that “the associations in our heads belong to us” (p. 553). This certainly does not mean that the cultural milieu has no impact on people's implicit attitudes. On the contrary, we assume that culture is a more powerful transmitter of opinions, values, beliefs, and ideologies (e.g., Devine, 1989; Jost & Banaji, 1994) than do those who wish to maintain a strict boundary between personal and “extrapersonal” (cultural) associations.

Tetlock and Mitchell see no contradiction between their claim (based exclusively upon public opinion data from white respondents) that prejudice no longer afflicts U.S. culture, on one hand, and their readiness to embrace others'

<sup>19</sup> Writing about attitudes in general, Asch (1952/1987) put it well: “The relation of social and personal processes has . . . posed a difficult problem. To cut through to a solution thinkers have often favored one of two extreme answers. Some have explained social events as the product of strictly individual tendencies (a favorite procedure of psychologists). More historically oriented students have tried to see the actions of individuals as a ‘reflection’ of social forces. . . . The study of attitudes may open a way to clarification of the problem. Here is a critical point at which social and personal processes join each other, a point at which social events become personally significant and personal events become of social moment” (p. 593; cf. Olson and Fazio, 2004).

<sup>20</sup> Rather than latching onto a potentially problematic dichotomy between personal and cultural origins of implicit attitudes and the associated critiques by Karpinski and Hilton (2001) and Olson and Fazio (2004), Tetlock would have been better off considering the implications for implicit attitudes of his own more perspicacious view of analogous dichotomies between public vs. private self-concepts and impression management vs. “purely” intrapersonal motivations (Tetlock and Levi, 1982; Tetlock and Manstead, 1985).

arguments that exposure to cultural biases is to “blame” for implicit bias, on the other. If we accept that (a) the culture is now free of bias, and (b) implicit associations “merely” reflect cultural beliefs, then it follows that (c) implicit measures should reveal little or no evidence of bias. But this conclusion is plainly false, as an empirical matter (e.g., Nosek et al., 2002; Nosek, Greenwald, et al., 2007; Nosek, Smyth, et al., 2007). So, which is it? Is the culture now free of prejudice, or is it the real perpetrator of implicit bias?

#### 3.2.4. Sympathy for the disadvantaged

Tetlock and Mitchell (2009) also propose that the IAT could be driven by sympathy, rather than antipathy, toward outgroup members. This argument is based on a single article by Uhlmann, Brescoll, and Paluck (2006) in which participants were classically conditioned to view two fictional groups in a status hierarchy, so that one group was seen as the victim of the other’s oppression. The words used to associate one group with victimization were: *oppressed*, *victimized*, *mistreated*, and *brutalized*; the words associated with the other group were: *privileged*, *rulers*, *dominant*, and *powerful*. Once trained, participants performed an IAT that contrasted the two groups with words representing “oppressed” and “privileged” and they also performed a traditional “good/bad” IAT. Scores on the two measures converged, and Tetlock and Mitchell now interpret this correlation as suggesting that the IAT “failed” to differentiate between groups that were seen as “badly off” and groups that were simply seen as “bad.”

However, because Uhlmann et al. (2006) used negatively valenced words to represent the “oppressed” category (e.g., *victim*, *oppressed*) and positively valenced words to represent the “privileged” category (e.g., *privileged*, *power*), it is quite plausible that the correlation with “good” vs. “bad” was due to similarity in valence. If the researchers had instead used positive words to represent the “oppressed” (e.g., *heroic*, *resistance*, *freedom-fighter* or even *humble* and *self-sacrificing*) and negative words to represent “privileged” (e.g., *bully*, *tyrant*, *oppressor*, *self-important*, *aloof*) they might not have observed the same correspondence between a (truly) sympathy-based IAT and a prejudice IAT. The original authors were careful to mention this methodological confound in their article (and also to raise the possibility that the correlation was due to a system-justifying tendency to blame victims for their own misfortune), but their qualifications were omitted in Tetlock and Mitchell’s retelling. Furthermore, if the IAT measured sympathy for the disadvantaged, then those with stronger anti-Arab IAT bias should be *more* rather than less likely to hire Arabs, and doctors with stronger anti-black IAT bias should treat black patients with *more* rather than less effusive treatment. As we have shown above, the opposite is true in both cases (Green et al., 2007; Rooth, 2007).

Finally, it should be noted that implicit evaluations of social groups frequently do correlate significantly with perceptions of their social or cultural status, and this seems to be true regardless of one’s own group membership (Rudman et al., 2002). Because social status is more easily discounted when people report their attitudes toward majority and minority groups at an explicit level, it is conceivable that measures of implicit bias are more sensitive to the psychological effects of hierarchical social systems. The observed pattern can be characterized broadly as “If group X is culturally valued (or possesses a great deal of symbolic and/or material resources), it is good,” and it corroborates predictions of system justification theory (Jost & Banaji, 1994; Jost et al., 2004; Rudman et al., 2002; Uhlmann et al., 2002). Thus, as with the parable of the “Two Jesses,” there is a clear theoretical interpretation of the key finding from the Uhlmann et al. (2006) study that reinforces rather than undermines the notion that the IAT taps a form of (system-justifying) bias rather than sympathy.

#### 3.2.5. Fear of being labeled a bigot, not bigotry

Rather than placing a reputational bet on or providing direct evidence for any single account of participants’ scores on measures of implicit bias, Tetlock and Mitchell throw several artifactual possibilities at the wall, hoping one of them will stick. In addition to the three we have already discussed, these critics suggest that IAT scores could simply reflect participants’ fears of being labeled a bigot rather than bigotry per se. This possibility, too, seems to spring from an (over-) interpretation of a single article (Frantz, Cuddy, Burnett, Ray, & Hart, 2004).

At first blush, it sounds impressive that when whites are told that the black–white IAT measures (personal) “racism” they perform “worse” (in a stronger pro-white direction) than when they are told that it measures “cultural” biases. However, a closer look at the findings of Frantz et al. (2004) reveals that neither of these two groups performed differently (on average) than did participants assigned to a control condition who were told nothing at all about the test. Thus, the fear of being labeled a bigot does not exert a sufficiently powerful effect on IAT scores to “explain away” implicit bias scores or the link between implicit bias scores and unfriendly behavior during cross-race interactions, as Tetlock and Mitchell conjecture. Furthermore, several studies of implicit bias have used *subliminal* primes (e.g.,

Dovidio et al., 1997), thereby rendering the “fear of bigotry” explanation even more farfetched. In any case, the fear of being labeled a bigot is hardly a viable alternative to the explanations given by implicit bias researchers, and it is misleading for critics to suggest that it is.

### 3.2.6. *Cognitive dexterity*

Tetlock and Mitchell also revive the early criticism of IAT research that it is merely a measure of cognitive skill or dexterity rather than group bias, but this hardly begins to make sense of the patterns of data produced by implicit bias researchers. For instance, would Tetlock and Mitchell assume that white conservatives possess a deficit in terms of cognitive skill or dexterity in comparison with, say, black liberals, based on the data summarized in Table 3? It is also important to point out—especially given that Tetlock and Mitchell do not—that the improved scoring method for the IAT reduces unwanted variance associated with individual differences in reaction-time considerably (Cai, Sriram, Greenwald, & McFarland, 2004; Greenwald, Nosek, et al., 2003; Greenwald, Oakes, et al., 2003; see also Mierke & Klauer, 2003). Nevertheless, with respect to the larger issue, it is necessary to keep in mind that individual differences in task performance can contribute to error variance (i.e., “statistical noise”) regardless of the instrument. For example, reading speed influences scores on intelligence tests, but it would be fairly absurd to conclude that intelligence tests are useless as a result of this influence.<sup>21</sup>

### 3.3. *Is organizational accountability a panacea?*

In dismissing the notion that implicit biases could derail good faith efforts to guarantee equal opportunity in today’s workplace, Tetlock and Mitchell (2009) are optimistic, if not Pollyannaish, about the notion that organizational accountability will solve the problem. They are deeply admiring of corporate efforts to curb bias (and we are, too, in some cases), but their chapter leads one to wonder why corporations would bother to take such steps, given that, according to Tetlock and Mitchell, it is practically self-evident that prejudice is no longer a problem in American society. We see several reasons to doubt that “accountability”—defined as “the implicit or explicit expectation that one may be called on to justify one’s beliefs, feelings, and actions to others” (Lerner & Tetlock, 1999, p. 251)—is a likely panacea for the problem of implicit bias.

As Lerner and Tetlock (1999) pointed out in their superb review of the research literature on accountability, “accountability is a logically complex construct that interacts with characteristics of decision-makers and properties of the task environment to produce an array of effects—only some of which are beneficial” (p. 270). That is, accountability often has no salutary effect on decision-making quality, especially when decision-makers are *unaware* of potentially distorting influences or lack the knowledge or training needed to recognize and overcome them. In fact, Lerner and Tetlock (1999) point out that increased accountability can even *exacerbate* some biases.

Research shows that it matters a great deal *to whom* a given decision-maker is held accountable. When superiors either tacitly or explicitly encourage bias, the results are likely to be disastrous for members of stigmatized minority groups (e.g., Brief et al., 2000). One fairly subtle way in which the encouragement to perpetuate biases might occur tacitly is when employers fail to “walk the talk,” that is, when they claim that they are in favor of equal opportunity, but their behavior suggests otherwise (see Pager & Quillian, 2005). When Pager (2007) conducted a review of several experimental field studies, including the study by Bertrand and Mullainathan (2004) summarized above, she found that whites were up to 5 times more likely to be interviewed and hired than were comparable black candidates. The managerial decision-makers in these studies were under precisely the kinds of “accountability pressures” trusted by Tetlock and Mitchell (2009) to curtail bias. Nevertheless, the decision-making outcomes were hardly egalitarian. Given such evidence, one wonders whether minority job candidates could ever take as much solace in accountability constraints as Tetlock and Mitchell propose.

<sup>21</sup> Tetlock and Mitchell (2009) also make much of Blanton and Jaccard’s (2006) argument that the IAT possesses an “arbitrary metric”, i.e., that absolute scores on the IAT are not directly interpretable. Once again, this is presented as a fatal flaw, but it is far from clear that metrics need to be non-arbitrary in order to be useful. To return to the example of intelligence testing (a domain in which scores are normed relative to others rather than linked to some absolute performance standard or benchmark), no one knows whether an IQ of 105 (or 115 or 125 or whatever) is adequate for someone to be an excellent business manager, lawyer, engineer, or fire fighter. Does this mean that intelligence is irrelevant to success in these occupations? Hardly. Does it mean that IQ tests are useless for measuring intelligence in the real world? We suspect not.



Moreover, even when managers report to superiors who are egalitarians in word and deed, accountability could easily translate into a relatively superficial concern with one's judgments and behaviors when decision-makers are forced to keep up with a fast-paced, harried environment that is characteristic of many organizations today. As Lerner and Tetlock (1999) put it:

Whether accountability does more than just alter public posturing depends on a host of moderators. . . The impression management processes triggered by accountability can interact in complex ways with cognitive processes. In this vein, one should not draw too sharp a distinction between private thought and public posturing (p. 269).

In offering organizational accountability as a possible panacea to the problem of implicit (as well as explicit) bias, Tetlock and Mitchell now seem to ignore all of the subtleties, complexities, and interaction effects that emerged from Lerner and Tetlock's (1999) review of the research on accountability. They have replaced that complexity with a simple—and, by Tetlock's own previous account, empirically untenable—main effect prediction: accountability will simply reduce (or even eliminate) bias.

Tetlock and Mitchell (2009) seek to discount findings from experimental studies of implicit bias by arguing that because research participants were not held accountable for their judgments and decisions, such laboratory findings cannot be generalized to the context of managerial decision-making. The authors assume that if participants were just held accountable for their decisions (as managers are), they would not be biased. Among other things, this assumption ignores several studies in which low-prejudice participants exhibited significant levels of implicit bias despite being informed that the task they were about to complete measured prejudice (e.g., Amodio et al., 2004, 2006, 2008).

One could also argue that participants in a research laboratory are under acute accountability pressures, given research on demand characteristics and social desirability biases (e.g., Orne, 1962; Paulhus, Bruce, & Trapnell, 1995). The fact that implicit prejudice is observed at all under typical laboratory conditions—when the average participant has plenty of access to cognitive resources, can often discern that the study is about racial (or other intergroup) attitudes, and is under significant impression management constraints—suggests that laboratory studies probably contribute to underestimates rather than overestimates of prejudice.<sup>22</sup> Moreover, a number of experiments have been conducted with actual managers as research participants (either in the laboratory or in on-line studies), and the results of these studies are highly similar to those obtained with college student participants (e.g., Bowles, Babcock, & Lai, 2007; Buttner & McEnally, 1996; Heilman, Block, & Martell, 1995; Martell, Parker, Emrich, & Crawford, 1998). For all of these reasons, we find ourselves in strong agreement with the sage recommendation of Lerner and Tetlock (1999), who cautioned their readers not to fall for “the falsity of the conventional wisdom—often born out of frustration at irrational, insensitive, or lazy decision-makers—that accountability is a cognitive or social panacea: ‘All we need to do is hold the rascals accountable’” (p. 270).

#### 4. Concluding remarks

Although we disagree strenuously with most of their arguments and the ways in which they have characterized research on (and, in some cases, researchers of) implicit bias, we are grateful to Tetlock and Mitchell (2009) for taking this general line of work seriously enough to attack it. We also appreciate the opportunity afforded us by the editors of this publication to summarize the reasons why we count ourselves among the scientific majority who accept the existence of implicit bias and also to clarify what we regard as major misconceptions (or perhaps misrepresentations) concerning research in this area. In closing, we submit that there is no need for either “fear or loathing” with respect to research programs on implicit bias, although both types of responses have often accompanied social scientific investigations of prejudice (and probably always will). On the contrary, there are many important reasons to embrace the insights arising from work on implicit bias, and we hope to have identified at least a subset of these. We hope also to have aired some empirically informed responses to the most egregious criticisms that have been leveled repeatedly against implicit bias researchers by those who have worked as sideline critics rather than being directly involved with the assessment of implicit bias and its consequences.

<sup>22</sup> Consistent with this notion, Fazio et al. (1995) found that participants scored lower on the Modern Racism Scale when they completed it in a laboratory and handed it to an experimenter than when they filled it out in a large, crowded lecture hall.

In conclusion, it seems to us that a world in which individuals, including organizational leaders, recognize the need for “unconsciousness raising” with respect to the subtle, often unintended effects of stigma, stereotyping, and prejudice will be a better world—and not at all a “stultifying regime of political correctness,” as Tetlock and Mitchell (2009, p. 24) worry. The danger seems small that we—as citizens or as social scientists—would be requiring too much of ourselves (or our organizations) by taking seriously the hundreds (if not thousands) of studies on implicit bias. These studies testify to the enduring historical (and psychological) legacies of racism, sexism, and other forms of group-based inequality and they suggest, not so surprisingly after all, that these legacies have left behind at least *some* residue in our hearts and minds.

## Acknowledgments

We wish to thank David M. Amodio, Mahzarin R. Banaji, Arthur P. Brief, Samuel D. Gosling, György Hunyady, Brian A. Nosek, and Barry M. Staw for extremely helpful comments and advice on earlier drafts. Briehan Truesdell provided significant administrative support for the project as a whole, for which we are grateful. The article was written while the lead author was supported by a research award from the National Science Foundation.

## References

- Allport, G. W. (1941). Psychological service for civilian morale. *Journal of Consulting Psychology*, 5, 235–239.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus Books. [Abridged version, 1958].
- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, 91, 652–661.
- Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*.
- Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality & Social Psychology*, 84, 738–753.
- Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, 15, 88–93.
- Amodio, D. M., Kubota, J. T., Harmon-Jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs. external cues. *Social Cognitive and Affective Neuroscience*, 1, 26–36.
- Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., & Amadori, A. (2008). Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, 29, 369–387.
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “would Jesse Jackson ‘fail’ the Implicit Association Test?”. *Psychological Inquiry*, 15, 257–278.
- Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans’ implicit racial associations and their implications for intergroup judgment. *Social Cognition*, 21, 61–87.
- Asch, S. (1952/1987). *Social psychology*. Oxford: Oxford University Press.
- Ayres, I. (2001). *Pervasive prejudice? Unconventional evidence of race and gender discrimination*. Chicago: University of Chicago Press.
- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68, 181–198.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7, 136–141.
- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, 65, 272–281.
- Bargh, J. A. (1994). The four horseman of automaticity: Awareness, intention, efficiency and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic processes* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A., & Pratto, F. (1986). Individual construct accessibility and perceptual selection. *Journal of Experimental Social Psychology*, 22, 293–311.
- Bartlett, F. C. (1932/1964). *Remembering: A study in experimental and social psychology*. London: Cambridge University Press.
- Bazerman, M. H., & Moore, D. (2008). *Judgment in managerial decision making* (7th ed.). New York: Wiley.
- Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95, 94–98.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? *American Economic Review*, 94, 991–1013.
- Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition*, 18, 329–353.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261.

- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*, 1142–1163.
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.
- Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*, 192–212.
- 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.
- Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*, 319–322.
- Bogardus, E. S. (1925). Social distance and its origins. *Journal of Applied Sociology*, *9*, 216–226.
- Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, *103*, 84–103.
- Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., & Vaslow, J. B. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational Behavior and Human Decision Processes*, *81*, 72–97.
- Buttner, E. H., & McEnally, M. (1996). The interactive effect of influence tactic, applicant gender, and type of job on hiring recommendations. *Sex Roles*, *34*, 581–591.
- Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition*, *22*, 673–684.
- Cohen, J., & Agiesta, J. (2008, June 22). 3 in 10 Americans admit to race bias. *Washington Post*. Retrieved July 15, 2008 from <http://www.washingtonpost.com/wp-dyn/content/story/2008/06/22/ST2008062200049.html>.
- Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314–1329.
- Correll, J., Park, B., Judd, C., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, *92*, 1006–1023.
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality & Social Psychology*, *44*, 20–33.
- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, *17*, 143–168.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.
- Dasgupta, N., Greenwald, A. G., & Banaji, M. R. (2003). The first ontological challenge to the IAT: Attitude or mere familiarity? *Psychological Inquiry*, *14*, 238–243.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316–328.
- Dasgupta, N., & Rivera, L. M. (2006). From automatic anti-gay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of Personality and Social Psychology*, *91*, 268–280.
- DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Caidric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science*, *15*, 319–324.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Devine, P. G. (2005). Breaking the prejudice habit: Airport's "inner conflict" revisited. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 327–342). Malden, MA: Blackwell.
- Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 339–360). New York, NY: Guilford Press.
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*, 835–848.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22–37.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In Zanna, M. P. (Ed.), *Advances in experimental social psychology*. Vol. 36 (pp. 1–52). San Diego, CA: Elsevier.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *84*, 754–770.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*, 573–598.
- Ebbinghaus, H. (1885/1913). *Memory. A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, representation, and visual perception. *Journal of Personality and Social Psychology*, *87*, 876–893.
- Faigman, D. L., Dasgupta, N., & Ridgeway, C. L. (2008). A matter of fit: The law of discrimination and the science of implicit bias. *University of California Hastings Law Review Journal*, *59*, 1389–1434.

- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes. A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.
- 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*, 229–323.
- Fiske, S. T. (1998). Stereotyping, prejudice and discrimination. In Gilbert, D. T., Fiske, S. T., & Lindzey, G. Eds. *Handbook of social psychology*. Vol. 2 (pp.357–411). New York: McGraw-Hill.
- Fiske, S. T. (2003). The discomfort index: How to spot a really good idea whose time has come. *Psychological Inquiry*, *14*, 201–206.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1991). Social science research on trial. *American Psychologist*, *46*, 1049–1060.
- Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*, 1611–1624.
- French, R. M., & Cleeremans, A. (2002). *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making*. New York: Psychology Press.
- Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, *46*, 23–30.
- Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, *44*, 164–172.
- Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology*, *114*, 304–308.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L., et al. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, *22*, 1231–1238.
- 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., & Banaji, M. R. (2003a). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.
- Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006a). Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard. *American Psychologist*, *61*, 56–61.
- Greenwald, A. G., Oakes, M. A., & Hoffman, H. (2003b). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, *39*, 399–405.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Measuring and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.
- Greenwald, A. G., Rudman, L. A., Nosek, B. A., & Zayas, V. (2006b). Why so little faith? A reply to Blanton and Jaccard's (2006) skeptical view of testing pure multiplicative theories. *Psychological Review*, *113*, 170–180.
- Gross, E. F., & Hardin, C. D. (2007). Implicit and explicit stereotyping of adolescents. *Social Justice Research*, *20*, 140–160.
- Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, *65*, 1105–1118.
- Hardin, C. D., & Banaji, M. R. (in press). The nature of implicit prejudice: Implications for personal and public policy. In E. Shafir (Ed.), *Policy implications of behavioral research*.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356–388.
- Hebb, D. O. (1949). *Organization of behavior*. Wiley: New York.
- Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality*, *10*, 237–252.
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, *89*, 416–427.
- Herman, E. (1995). *The romance of American psychology: Political culture in the age of experts*. Berkeley: University of California Press. Available online: <http://ark.cdlib.org/ark:/13030/ft696nb3n8/>.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.
- 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.
- Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes and Intergroup Relations*, *11*, 69–87.
- Jackman, M. R. (1994). *The velvet glove: Paternalism and conflict in gender, class and race relations*. Berkeley, CA: University of California Press.
- Jolls, C., & Sunstein, C. R. (2006). The law of implicit bias. *California Law Review*, *94*, 969–996.
- Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system—Justification and the production of false—Consciousness. *British Journal of Social Psychology*, *33*, 1–27.
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–919.
- Jost, J. T., & Hamilton, D. L. (2005). Stereotypes in our culture. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 208–224). Malden, MA: Blackwell.
- Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509.

- Jost, J. T., Pelham, B. W., & Carvallo, M. (2002). Non-conscious forms of system justification: Cognitive, affective, and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*, 586–602.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kang, J. (2005). Trojan horses of race. *Harvard Law Review*, *118*, 1489–1593.
- Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of “affirmative action”. *California Law Review*, *94*, 1063–1118.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.
- Katz, D., & Braly, K. W. (1935). Racial prejudice and racial stereotypes. *Journal of Abnormal and Social Psychology*, *30*, 175–193.
- Krieger, L. H., & Fiske, S. T. (2006). Behavioral realism in employment discrimination law: Implicit bias and disparate treatment. *California Law Review*, *94*, 997–1062.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: Procedures and validity. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59–102). New York: Guilford Press.
- LeDoux, J. E. (1996). *The emotional brain*. New York: Simon and Schuster.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092–1107.
- Lewin, K. (1941). Self-hatred among Jews. *Contemporary Jewish Record*, *4*, 219–282.
- Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans’ implicit and explicit racial attitudes. *Journal of Experimental Social Psychology*, *38*, 405–413.
- Locke, E. A. (Ed.). (1986). *Generalizing from laboratory to field settings*. Lexington, MA: Lexington Books.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.
- MacIntyre, A. C. (1958). *The unconscious: A conceptual study*. London: Routledge & Kegan Paul.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808–817.
- March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.
- Martell, R. F., Parker, C., Emrich, C. G., & Crawford, M. S. (1998). Sex stereotyping in the executive suite: “Much ado about something”. *Journal of Social Behavior & Personality*, *13*, 127–138.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, explicit attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology*, *37*, 435–442.
- McConnell, A. R., & Leibold, J. M. (2009). Weak evidence and selective criticisms: Reply to Blanton et al.. *Journal of Applied Psychology*, *94*, 583–589.
- Mendes, W. B., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. S. (2007). Why egalitarianism might be good for your health: Physiological thriving during stressful intergroup encounters. *Psychological Science*, *18*, 991–998.
- Mentovich, A., & Jost, J. T. (2008). The ideological “id”? System justification and the unconscious perpetuation of inequality. *Connecticut Law Review*, *40*, 1095–1116.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.
- Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, *192*(4234), 27–33.
- Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*, 1180–1192.
- Mitchell, G., & Tetlock, P. E. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State University Law Review*, *67*, 1023–1122.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*, 655–663.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality & Social Psychology*, *83*, 1029–1050.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379–387.
- Moskowitz, G. B., Gollwitzer, P., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*, 167–184.
- Myrdal, G. (1944). *An American dilemma: The Negro problem and modern democracy*. New York: Harper.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.
- New Hampshire v. Michael Addison*, 997 A. 520. (2009).
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Nock, M. K., & Banaji, M. R. (2007). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Clinical and Consulting Psychology*, *75*, 707–715.
- Nosek, B. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *132*, 565–584.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.

- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007a). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York, NY: Psychology Press.
- Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*, 553–594.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait–multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*, 14–29.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007b). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.
- Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.
- Ostafin, B. D., & Palfai, T. P. (2006). Compelled to consume: The Implicit Association Test and automatic alcohol motivation. *Psychology of Addictive Behaviors*, *20*, 322–327.
- Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: The effect of word familiarity and frequency on the Implicit Association Test. *Social Cognition*, *19*, 97–144.
- Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology*, *108*, 937–975.
- Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *Annals of the American Academy of Political & Social Science*, *609*, 104–133.
- Pager, D., & Quillian, L. (2005). Walking the talk: What employers say versus what they do. *American Sociological Review*, *70*, 355–380.
- Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: Assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy*, *41*, 1149–1162.
- Park, S. H., Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice moderates the effect of cognitive depletion on unintended discrimination. *Social Cognition*, *26*, 401–419.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, *21*, 100–108.
- Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, *26*, 199–216.
- Pfeffer, J. (1998). Understanding organizations: Concepts and controversies. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 733–777). New York: Oxford.
- Pfeffer, J., Cialdini, R. B., Hanna, B., & Knopoff, K. (1998). Faith in supervision and the self enhancement bias: Two psychological reasons why managers don't empower workers. *Basic and Applied Social Psychology*, *20*, 313–321.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlling processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Gatenby, J. C., Funayama, E. S., Gore, J. C., et al. (2000). Amygdala activation predicts performance on indirect measures of racial bias. *Journal of Cognitive Neuroscience*, *12*, 729–738.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832.
- Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*, *37*, 486–501.
- Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *16*, 180–183.
- Price-Waterhouse v. Hopkins*, 109 S. Ct. 1775. (1989).
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, *11*, 37–43.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385–408.
- Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading-activation theories versus compound-cue theories. *Psychological Review*, *101*, 177–184.
- Riach, P. A., & Rich, J. (2002). Field experiments of discrimination in the market place. *Economic Journal*, *112*, F480–F518.
- Richeson, J. A., & Shelton, J. N. (2005). Brief report: Thin slices of racial bias. *Journal of Nonverbal Behavior*, *29*, 75–86.
- Richeson, J. A., & Trawalter, S. (2005). Why do interracial interactions impair executive function? A resource depletion account. *Journal of Personality and Social Psychology*, *88*, 934–947.
- Roberson, L., Kulik, C. T., & Pepper, M. B. (2001). Designing effective diversity training: Influence of group composition and trainee experience. *Journal of Organizational Behavior*, *22*, 871–885.
- Roberson, L., Kulik, C. T., & Pepper, M. B. (2003). Using needs assessment to resolve controversies in diversity training design. *Group and Organization Management*, *28*, 148–174.
- Rooth, D. (2007). *Implicit discrimination in hiring: Real world evidence* (IZA Discussion Paper No. 2764). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor).
- Rudman, L. A. (2004). Social justice in our minds, homes and society: The nature, causes and consequences of implicit bias. *Social Justice Research*, *17*, 129–142.
- Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the Implicit Association Test. *Group Processes and Intergroup Relations*, *10*, 359–372.
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). “Unlearning” automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology*, *81*, 856–868.
- Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology*, *31*, 493–517.

- Rudman, L. A., Feinberg, J. M., & Fairchild, K. (2002). Minority members' implicit attitudes: Ingroup bias as a function of group status. *Social Cognition, 20*, 294–320.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743–762.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17*, 1–29.
- Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes and Intergroup Relations, 5*, 133–150.
- Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations* (revised ed.). Cambridge, MA: Harvard University Press.
- Sears, D. O. (1994). Ideological bias in political psychology: The view from scientific hell. *Political Psychology, 15*, 547–556.
- Sears, D. O., van Laar, C., Carrillo, M., & Kosterman, R. (1997). Is it really racism? The origins of white Americans' opposition to race-targeted policies. *Public Opinion Quarterly, 61*, 16–53.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
- Sidanius, J., Pratto, E., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology, 70*, 476–490.
- Simon, H. A. (1957). *Models of man*. New York: Wiley.
- Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative orientation. *Journal of Personality and Social Psychology, 89*, 583–592.
- Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic Racism: Problems of motive attribution in political analysis. *Journal of Social Issues, 42*, 129–150.
- Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial response among aversive racists. *Journal of Experimental Social Psychology, 38*, 71–77.
- Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37*, 1660–1672.
- Stangor, C., Sullivan, L. A., & Ford, T. E. (1991). Affective and cognitive determinants of prejudice. *Social Cognition, 9*, 359–380.
- Staw, B. M. (1991). Dressing up like an organization: When psychological theories can explain organizational action. *Journal of Management, 17*, 805–819.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
- Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology, 15*, 509–529.
- Tetlock, P. E., & Arkes, H. R. (2004). The implicit prejudice exchange: Islands of consensus in a sea of controversy. *Psychological Inquiry, 15*, 311–321.
- Tetlock, P. E., & Levi, A. (1982). Attribution bias: On the inconclusiveness of the cognition–motivation debate. *Journal of Experimental Social Psychology, 18*, 68–88.
- Tetlock, P. E., & Manstead, A. S. R. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review, 92*, 59–77.
- Tetlock, P. E., & Mitchell, G. (2008). Calibrating prejudice in milliseconds. *Social Psychology Quarterly, 71*, 12–16.
- Tetlock, P. E., & Mitchell, G. (2009). Implicit Bias and Accountability Systems: What Must Organizations do to Prevent Discrimination? *Research in Organizational Behavior, 29*, 3–38.
- Thush, C., & Wiers, R. W. (2007). Explicit and implicit alcohol-related cognitions and the prediction of future drinking in adolescents. *Addictive Behaviors, 32*, 1367–1383.
- Thush, C., Wiers, R. W., Ames, S. L., Grenard, J. L., Sussman, S., & Stacy, A. W. (2007). Apples and oranges? Comparing indirect measures of alcohol-related cognition predicting alcohol use in at-risk adolescents. *Psychology of Addictive Behaviors, 21*, 587–591.
- Towles-Schwen, T., & Fazio, R. H. (2006). Automatically activated racial attitudes as predictors of the success of interracial roommate relationships. *Journal of Experimental Social Psychology, 42*, 698–705.
- Tyler, T. R., Boeckmann, R., Smith, H. J., & Huo, Y. (1997). *Social justice in a diverse society*. Boulder, CO: Westview Press.
- Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A., & Swanson, J. (2002). Subgroup prejudice based on skin color among Hispanics in the United States and Latin America. *Social Cognition, 20*, 198–226.
- Uhlmann, E. L., Brescoll, V. L., & Paluck, E. L. (2006). Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *Journal of Experimental Social Psychology, 42*, 491–499.
- von Hippel, W., Brener, L., & von Hippel, C. (2008). Implicit prejudice toward injecting drug users predicts intentions to change jobs among drug and alcohol nurses. *Psychological Science, 19*, 7–11.
- Wax, A. (2008). The discriminating mind: Define it, prove it. *Connecticut Law Review, 40*(4), 1–41.
- Wax, A., & Tetlock, P. E. (2005, December 1). We are all racists at heart. *Wall Street Journal*, A16.
- Wiers, R. W., Houben, K., & de Kraker, J. (2007). Implicit cocaine associations in active cocaine users and controls. *Addictive Behaviors, 32*, 1284–1289.
- Wiers, R. W., Van Woerden, N., Smulders, F. T. Y., & De Jong, P. J. (2002). Implicit and explicit alcohol related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111*, 648–658.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10*, 109–120.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, Monographs 9*(2 Pt 2).
- Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology, 90*, 553–562.