

# The IAT Is Dead, Long Live the IAT: Context-Sensitive Measures of Implicit Attitudes Are Indispensable to Social and Political Psychology

Current Directions in Psychological Science

1–10

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0963721418797309

www.psychologicalscience.org/CDPS



**John T. Jost** 

Department of Psychology, New York University

## Abstract

The implicit association test (IAT) is one of several measures of implicit attitudes, but it has attracted especially intense criticism. Some methodological objections are valid, but they are damning only if one accepts false analogies between the IAT and measures of intellectual aptitude, clinical diagnosis, or physical height. Other objections are predicated on misconceptions of the nature of attitudes (which are context-sensitive and reflect personal and cultural forces) or the naive assumption that people cannot be biased against their own group. Other criticisms are ideological, pertaining to questions of moral and political value, such as whether it is good to have fewer pro-White/anti-Black implicit attitudes and to provide respondents with feedback about their implicit attitudes. Implicit-attitude measures have been extremely useful in predicting voting and other political behavior. An indirect, unobtrusive, context-sensitive measure of attitudes is far more useful to social and political psychologists than an IQ test or clinical “diagnosis” would be, insofar as it reflects a dynamic Lewinian conception of the “person in the situation.”

## Keywords

attitudes, implicit social cognition, bias, prejudice, IAT

There is a cluster of methodological and ideological objections to research on implicit bias—especially research using the implicit association test (IAT)—that has circulated for more than a decade (Arkes & Tetlock, 2004; Blanton & Jaccard, 2006; Mitchell & Tetlock, 2017; Tetlock & Mitchell, 2009). Because these objections have been repeated so many times and receive prominence in journalistic critiques—including articles in *New York* magazine and *The Chronicle of Higher Education* (Bartlett, 2017; Singal, 2017)—some people may conclude that these critiques are damning and that researchers of implicit bias have simply ignored them. Both conclusions would be false (e.g., Jost et al., 2009a, 2009b).

## Rebutting Methodological Objections

Major psychometric critiques are that the IAT exhibits modest test-retest reliability (Bar-Anan & Nosek, 2014; Gawronski, Morrison, Phillips, & Galdi, 2017), that social-contextual factors affect IAT scores (Barden, Maddux,

Petty, & Brewer, 2004; Dasgupta & Greenwald, 2001; Lowery, Hardin, & Sinclair, 2001), and that there are no clear cutoff points between being biased and unbiased (Blanton & Jaccard, 2006; Mitchell & Tetlock, 2017). These observations are valid for the most part, but to wield them as critiques, it is necessary to create false or misleading analogies between the IAT and measures of performance or aptitude (IQ tests), measures of clinical diagnosis (tests of depression), or physical properties that are assumed to be invariant. In an especially absurd comparison, the IAT was likened to measuring height with a stack of melting ice cubes (Singal, 2017).

The IAT is none of these things; it is one of several indirect behavioral measures of an individual's attitudes (see Fig. 1). Social psychologists, at least, ought to realize that attitudes, which may be defined as “object-evaluation

## Corresponding Author:

John T. Jost, New York University, Department of Psychology, 6 Washington Place, Room 610, New York, NY 10003  
 E-mail: john.jost@nyu.edu

Task step	1	2	3	4	5
Task Description	Initial Target-Concept Discrimination	Associated Attribute Discrimination	Initial Combined Task	Reversed Target-Concept Discrimination	Reversed Combined Task
Task Categories	MALE • FEMALE •	career • family •	MALE • career • FEMALE • family •	MALE • FEMALE •	MALE • career • FEMALE • family •
Stimuli	o TERESA o HELEN o JOHN o JANET o SAMUEL o MIKE o DINAH o THOMAS	o office o executive o kitchen o parents o manager o family o salary o home	o salary o THOMAS o kitchen o MARY o executive o DAVID o babies o HELEN	MARY o JOHN o THOMAS o DINAH o SAMUEL o HELEN o TERESA o MIKE o	o DAVID o career o SARAH o wedding o JANET o employment o HARRY o babies

**Fig. 1.** Schematic illustration of the implicit association test (IAT). The IAT involves a series of discrimination tasks (numbered columns). A pair of target concepts and an attribute dimension are introduced in the first two steps. Categories for each of these discriminations are assigned to a left-hand or right-hand response on a computer keyboard, indicated here by the black circles. For example, participants are first shown a series of names such as “TERESA” and asked to categorize them as “MALE” or “FEMALE”; in the next step, they may be asked to categorize a series of words as related to either “career” or “family.” These two categories (and associated stimuli) are combined in the third step and then recombined in the fifth step after response assignments are reversed (in the fourth step) for the target-concept discrimination. Correct responses in this example are indicated by open circles. (Adapted from Greenwald, McGhee, & Schwartz, 1998.)

associations of varying strength” (Fazio, 2007, p. 619), can and do change—even if they are sometimes resistant to change—and their expression is highly “influenced by salient contextual information” (Fazio, 2007, p. 619). Consequently, there is no sense in which modest test-retest reliability or demonstrations that IAT scores are affected by social-contextual factors—such as demographic characteristics of experimenters (e.g., Lowery et al., 2001)—can be said to provide evidence against the existence of an attitude as long as our theory of attitudes assumes some degree of contextual variability.

In looking back, Banaji (2004) acknowledged that she at first assumed that because “implicit attitudes were automatic and relatively uncontrollable,” it followed that “environmental probes could not shape or shift them” (p. 138). However, she changed her view on the basis of empirical demonstrations of contextual sensitivity (e.g., Dasgupta & Greenwald, 2001; Lowery et al., 2001) and concluded that implicit attitudes are (sometimes) “elastic in their response to even subtle suggestions in the environment” (pp. 139–140; see also Hardin & Banaji, 2013). In any case, the fact remains that the IAT exhibits higher (within-persons) test-retest reliability than other response-latency measures commonly used in psychological research, including Stroop and priming tasks (Bar-Anan & Nosek, 2014). Like the Stroop task, the IAT is highly reliable at the aggregate

level in that it consistently produces very similar means from sample to sample within the same population.

In addition, the psychometric properties of the standard IAT have been found to be superior to many other measures of implicit attitudes, including the go/no-go association task, brief IAT, single-category IAT, personalized IAT, and pencil-and-paper IAT (Bar-Anan & Nosek, 2014; Kurdi et al., 2018). As in the case of explicit attitudes (e.g., Fazio, 2007), there is some stability and some malleability when it comes to implicit attitudes—and measures of implicit attitudes such as the IAT capture both elements. If one insists on drawing an analogy to physical measurement, IAT scores are much more like blood-pressure readings than height estimates. They can be affected by situational factors (much as blood-pressure readings are affected by exertion or caffeine), but multiple tests of the same individual are likely to yield reasonable generalizations about that individual. In fact, the test-retest reliability for many types of IATs is as high as (or higher than) that for self-administered blood-pressure readings (e.g., see Brody, Veit, & Rau, 1999).

It is true that there may be no precise “zero point” between liking and disliking an attitude object (especially insofar as many attitudes are ambivalent and multidimensional), but it does not follow that there is no conceptual difference between liking and disliking or

that relative (vs. absolute) estimates of current preferences are useless. (The cutoff between “low” and “high” blood pressure, likewise, is a matter of professional convention, and it will always be at least somewhat arbitrary.) Blanton and Jaccard (2006) acknowledged that “the issue of metric arbitrariness is irrelevant for many research goals” and that if one “wishes to test if variables pattern themselves in ways predicted by a theory, then there usually will be no need to conduct studies to reduce the arbitrariness of the metric” (p. 39). If the IAT is properly understood as a context-sensitive measure of implicit attitudes—rather than something resembling an IQ test or clinical diagnosis—then the demand for a “nonarbitrary metric” disappears. Relative (implicit) preferences are indeed useful in many areas of social and political psychology for testing theoretically derived predictions about, among other things, candidate preferences, ideological inclinations, degree of trust in government, and the internalization of aspects of social, economic, and political systems.

It is important to revisit questions of predictive validity arising from meta-analyses reporting that average correlations between race-based IATs and discriminatory behaviors ranged from approximately .13 to .24 (Carlsson & Agerström, 2016; Greenwald, Banaji, & Nosek, 2015; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Authors of another review suggested that interventions designed to change implicit bias did not necessarily translate into changes in behavior (Forscher et al., 2016), a finding that has been interpreted as proof that the IAT has been overhyped (Singal, 2017). But there are probably several reasons why past summaries have turned up fairly low correlations between implicit racial attitudes and behavioral outcomes, including the fact that (a) in many studies, measures of implicit attitudes and behaviors were low in methodological correspondence, and (b) researchers have seldom adjusted properly for measurement error (Greenwald et al., 2015; Greenwald et al., 2009; Kurdi et al., 2018). Furthermore, the interventions discussed by Forscher and colleagues (2016) were brief and minimalistic; it is not too surprising that they produced little in the way of lasting change.

When implicit and explicit attitudes are highly correlated—as in the case of many personal and political preferences, such as those involving candidates, parties, and ideological positions (Nosek, 2005, p. 572)—implicit attitudes are reasonably strong predictors of behavior (Bos, Sheets, & Boomgaarden, 2018; Greenwald et al., 2009; Lundberg & Payne, 2014). A forthcoming meta-analysis by Kurdi et al. (2018)—which is based on a total sample size ( $N = 36,071$ ) that is 6 to 10 times larger than those used in previous meta-analyses by Greenwald et al. (2009) and Oswald et al. (2015)—is

not restricted to the topics of race and ethnicity, as the earlier reviews were. It suggests that standard IAT scores are indeed robust predictors of behavioral outcomes (with correlations as high, in some cases, as .37) and that they exhibit incremental validity (after adjusting for explicit attitudes in structural equation models)—especially when one focuses on high-quality studies using standard (as opposed to modified) IATs with large sample sizes. No evidence of publication bias was obtained (Kurdi et al., 2018).

It is well known in social psychology that a lack of measurement correspondence (or specificity) between attitudinal and behavioral measures undermines predictive validity (Ajzen & Fishbein, 1980). One solution that has proven useful in the case of explicit attitudes is to measure behavioral intentions (rather than more general or abstract attitudes). A similar approach may help to increase the predictive validity of implicit attitudes. For instance, Palfai and Ostafin (2003) modified the IAT procedure to measure implicit behavioral orientations (such as *approach* and *avoidance*) in a sample of problem drinkers presented with alcohol cues (namely a glass of beer placed in front of them)—thereby directly incorporating contextual factors into the study. Results revealed that people who exhibited stronger alcohol-approach associations anticipated more pleasure and satisfaction as they pondered the beer before them ( $r = .41, p < .01$ ) and felt a stronger urge to drink it ( $r = .30, p < .05$ ). They also admitted to more binge drinking over the previous month ( $r = .34, p < .05$ ). Procedures such as this one may be especially effective when it comes to appetitive attitudes (and behaviors), but focusing on behavioral intentions—such as approach and avoidance—might be one useful way of increasing attitude-behavior correspondence in the study of implicit social cognition.

This is potentially important, because Kurdi et al. (2018) systematically compared implicit attitude-behavior correlations for situations in which there was (according to multiple raters) high versus low correspondence (or measurement specificity) between implicit attitudes and behaviors. Examples of low-correspondence cases included those in which responses to a Black/White-pleasant/unpleasant IAT were used to predict global self-esteem, responses to a homosexual/heterosexual-pleasant/unpleasant IAT were used to predict drives for thinness and muscularity, and responses to an Arab/American-pleasant/unpleasant IAT were used to predict participants' ratings of male and female American targets as sexually aroused. Examples of high-correspondence cases included those in which responses to a fat/slim-good/bad IAT were used to predict participants' willingness to interact with an overweight person, responses to a people-with-AIDS/people-without-AIDS-pleasant/

unpleasant IAT were used to predict avoidance of people with AIDS, and responses to an Arab American/White–dangerous/safe IAT were used to predict the endorsement of racial profiling in airports. Although their results were somewhat complicated because coders differed in their ratings of correspondence as a function of familiarity with the study results, Kurdi and colleagues observed that correlations between implicit attitudes and behavior were much stronger—indeed, they were as strong as correlations between explicit attitudes and behavior—when measurement correspondence was high. When correspondence was low, however, implicit attitude–behavior correlations were small, albeit still statistically significant (see also Greenwald et al., 2009).

Some researchers have criticized the use of differences scores in response latencies to estimate individual differences in general, in part because of speed/accuracy trade-offs, and others have taken aim at the IAT in particular (Blanton, Jaccard, Gonzales, & Christie, 2006). It is important to keep in mind, however, that the IAT scoring algorithm takes into account accuracy as well as speed, and it also adjusts for participant-level variability in response latencies, thereby compensating for two major limitations of these kinds of differences scores. Nosek and Sriram (2007) addressed Blanton and colleague’s (2006) criticisms in great detail, arguing that they are derived from the faulty assumption that IAT responses to the same stimuli in different conditions (e.g., pleasant vs. unpleasant matching tasks) represent multiple items from the same “scale.”

### **Correcting Theoretical and Metatheoretical Misconceptions**

Additional objections to implicit-bias research are predicated on misunderstandings (or misrepresentations) of other foundational assumptions of social psychology. It is often claimed that the IAT simply measures familiarity with or awareness of cultural stereotypes rather than personal animus (Arkes & Tetlock, 2004; Karpinski & Hilton, 2001; Mitchell & Tetlock, 2017; Tetlock & Mitchell, 2009). Not only have researchers of implicit bias addressed these possibilities empirically (e.g., Dasgupta, McGhee, Greenwald, & Banaji, 2001; Nosek & Hansen, 2008; Ottaway, Hayden, & Oakes, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999), but also the question of whether implicit attitudes reflect personal preferences as opposed to social and cultural processes is ill-posed. (We also know from decades of research on the “mere-exposure effect” that familiarity breeds liking, so there is no reason to assume that familiarity and attitudinal evaluation should be unrelated.) In any case, such a strong juxtaposition between the personal and the social would have made little

sense to Kurt Lewin, Gordon Allport, and other pioneers of social psychology. For instance, Solomon Asch (1952/1987) wrote that

some [researchers] have explained social events as the product of strictly individual tendencies [whereas 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)

The fact is that implicit-bias researchers have long argued that it is possible for people to be biased without feeling personal animus and that exposure to shared cultural stereotypes (and objective inequalities among groups in the social system) are major sources of implicit bias (Banaji & Greenwald, 2013; Hardin & Banaji, 2013; Jost, Banaji, & Nosek, 2004; Jost et al., 2009a; Payne, Vuletich, & Lundberg, 2017). In responding directly to Karpinski and Hilton’s (2001) conclusion that IAT scores “reveal little about a person’s beliefs and much about his or her environment or culture” (p. 787), Banaji (2001) pointed out that it is a fallacy to assume that “there is a bright line separating one’s self from one’s culture, an assumption that is becoming less tenable as researchers discover the deep reach of culture into individual minds” (p. 139). She asked, rhetorically, whether anyone would claim that an Indian person’s love of spicy pickles is not a “genuine attitude” simply because its origin has a very strong cultural basis. Likewise, Nosek and Hansen (2008) concluded that “the associations in our heads belong to us” (p. 553) on the basis of evidence that individual differences in IAT scores were more strongly correlated with people’s explicit attitudes than with mere awareness of cultural evaluations of the same attitude objects.

An even weaker criticism, also still in circulation, is that the IAT does not measure intergroup bias because members of disadvantaged groups sometimes exhibit implicit preferences for out-groups—and people obviously cannot be biased (or prejudiced) against members of their own group (Arkes & Tetlock, 2004). This issue was dramatized by Tetlock and Mitchell (2009) in terms of “The Parable of the Two Jesses,” who were stipulated to hold the same implicit attitudes about race:

One Jesse (Jackson) . . . 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. (p. 12)

The problem with this argument is that it ignores 75 years of social science, including the work of Lewin, who analyzed the problem of “group self-hatred,” and the doll studies carried out by Kenneth and Mamie Clark.<sup>1</sup> This phenomenon has been investigated extensively by system-justification researchers, who observed that 40% or 50% of disadvantaged group members implicitly favor advantaged out-groups (Jost et al., 2004).

The research shows that poor people often favor rich people; obese people favor normal-weight people; Hispanics favor Whites, and dark-skinned Morenos favor light-skinned Blancos in Chile; Black children and children of color favor Whites in South Africa; and Blacks, Latinos, and Asians favor Whites in the United States. Field studies show that Black and Latino passengers tip White taxi drivers better than Black and Latino taxi drivers (Greenwald & Pettigrew, 2014), and gay people often favor straight people on implicit measures (Hoffarth & Jost, 2017; Jost et al., 2004). Some researchers contend that implicit-attitude measures fail to pick up anything useful that is not already captured by measures of explicit attitudes, but this is not the case with respect to out-group favoritism. It is relatively common for members of disadvantaged groups to favor their own group on explicit measures—perhaps because they do not want to see themselves as (or be seen as) an “Uncle Tom”—but to show out-group favoritism on implicit measures anyway (Jost et al., 2004).<sup>2</sup>

To answer Arkes and Tetlock’s (2004) initial question about “The Two Jesses,” Jesse Jackson probably would not “fail” the IAT, because Black liberals exhibit no implicit bias (on average). But Jesse Helms probably would “fail,” so to speak, because White conservatives exhibit a very strong pro-White/anti-Black bias (on average). Jost et al. (2004) observed that the effect size for the difference between these two groups is very large ( $d = 1.31$ ). The interaction between political ideology and group membership on implicit in-group favoritism is even clearer and stronger when it comes to sexual orientation. Straight conservatives show significantly stronger in-group favoritism than straight liberals, and gay conservatives show significantly stronger out-group favoritism than gay liberals (Jost et al., 2004). Does this mean that people who exhibit antigay implicit attitudes on one occasion are hopelessly and forever doomed to live gay-bashing lifestyles or to internalize

homonegativity? No, because attitudes can and do change, in part because they reflect a dynamic social and cultural environment, which can be harsh or welcoming for various groups of people at different times (e.g., Charlesworth & Banaji, 2018).

Payne et al. (2017) advanced the argument, which is consistent with system-justification theory, that much of the stability in aggregate levels of implicit bias is attributable to the stability of situations—the maintenance of social, cultural, and political institutions and arrangements. The smart money is on a Lewinian “person-in-the-situation” view (Ross & Nisbett, 2011), namely that measures of implicit attitudes reflect something important about the individual, including his or her ideological worldview, as well as something important about the surrounding social and cultural context, as Banaji (2001) and many other researchers pointed out years ago.

## Addressing Ethical and Ideological Objections

Once the methodological criticisms noted above are understood properly as relying on misleading analogies and faulty assumptions, the remaining objections can be seen more clearly as pertaining to questions of value (or politics), such as whether it is better to have fewer pro-White/anti-Black implicit attitudes and whether it is good or bad to give people feedback about their implicit attitudes (Blanton & Jaccard, 2006; French, 2017; Singal, 2017). These objections are to educational content presented on the Project Implicit website (<https://implicit.harvard.edu/>), or to the fact that Anthony Greenwald and Mahzarin Banaji called a press conference in 1998, or that they and their students have advised lawyers, judges, and business executives. In enumerating these alleged transgressions, Mitchell and Tetlock (2017) lamented “the ideological sympathy that many social psychologists likely have for the implicit prejudice meme” and worried that “the road to reduce the popularity of the IAT through ordinary science looks long and difficult to navigate” (p. 187).

It is to be expected that people will differ in the value judgments they attach to questions of implicit and explicit bias. But if the IAT measures a particular attitude at a given time in a specific social context, there is nothing inherently problematic (or unethical) about providing people with feedback concerning that attitude measurement, recognizing that it is, after all, only one measurement at one point in time. Nevertheless, some people might dislike (for ideological reasons, including motivated system defensiveness) the notion that such feedback might be interpreted in light of critical sociological theory—or that it might be used to

generate broader discussions about prejudice in society (French, 2017; Mitchell & Tetlock, 2017; Singal, 2017). But surely it is preferable to live in a society in which scientists are encouraged to speak freely about what they see as the social and political implications of their research, even if some of those implications are debatable, than to live in a society in which those individuals are attacked and accused of professional misconduct for doing so.

There are clearly legitimate (and important) ethical questions about how the IAT and other implicit measures—including methods commonly used in neuroscience research—should and should not be used, in light of privacy and other concerns (Richmond, Rees, & Edwards, 2012). For instance, it may be tempting for law-enforcement officials to use implicit measures to monitor the likelihood of recidivism among convicted sexual predators (Nunes, Firestone, & Baldwin, 2007). Such efforts should not (and presumably would not) proceed without extensive moral and legal deliberation, for there are clear and obvious dangers associated with using “mind-reading” techniques to incarcerate or hospitalize people against their will.

Researchers of implicit attitudes, in any case, have warned that there are profound ethical concerns about holding someone accountable for his or her attitudes (whether implicit or explicit)—as opposed to their deeds. The Project Implicit website warns against using the IAT to “diagnose an individual” or to “choose jurors,” as practical examples, while pointing out that “it might be appropriate to use the IAT to teach jurors about the possibility of unintended bias” (Project Implicit, 2011). Likewise, as Kurdi et al. (2018) wrote,

given such malleability, we have always advised against using a single intergroup IAT as a device for the selection of people, such as whether to hire someone for a job or admit them to a club. The measure is of value in two contexts: research and education. (p. 37)

It makes little sense to criticize basic research on implicit attitudes on the grounds that it could, one day, be misused—especially when researchers have consistently argued against such applications.

As a “bona fide” pipeline used to quantify levels of “unconscious racism” as a fixed property of the individual—or as a diagnostic tool to classify people as “having” racism or sexism (like they might “have” clinical depression)—the IAT is dead. I do not know if any researchers of implicit bias actually conceived of the IAT in these ways, but critics continue to assert that this is our conception (Bartlett, 2017; Mitchell & Tetlock, 2017; Singal, 2017). It certainly is not mine.

Nor do I know of any researchers who were so naive as to assume that the IAT tells “a simple, pat story about how racism works and can be fixed: that deep down, we’re all a little—or a lot—racist, and that if we measure and study this individual-level racism enough, progress toward equality will ensue” (Singal, 2017, para. 9). On the contrary, Hardin and Banaji (2013) pointed out that “culture-wide changes in implicit prejudice will require culture-wide changes in social organization and practice” and that work on “implicit prejudice is likely to be either encouraging or depressing, depending upon one’s sense of the likelihood of broad, long-term changes in social organization and culture” (p. 21). Charlesworth and Banaji (2018) found that implicit attitudes have become more egalitarian over the last decade when it comes to sexual orientation, race, and skin-tone—but not age, disability, or obesity.

The IAT as a “magic bullet”—a panacea for solving the world’s ongoing problems with racism and sexism and classism—is dead. Nevertheless, social and political psychologists still need good methods for assessing, indirectly and unobtrusively, social and political attitudes in a wide range of contexts. It is reasonable to begin with existing methods and make them better, and that is what most researchers of implicit bias have been doing for years. Until an obviously superior technology for measuring implicit attitudes is devised, researchers are justified in employing the IAT—for it continues to outperform the available alternatives in terms of reliability and validity (Bar-Anan & Nosek, 2014; Kurdi et al., 2018). Rather than dismissing the study of implicit bias altogether, it would be more constructive for methodologically qualified critics to join ongoing efforts to improve its measurement (Jost et al., 2009b).

### **A Brief Review of Evidence That Measures of Implicit Attitudes Illuminate Political Psychological Phenomena**

The IAT is not a pristine or “process-pure” measure of attitudes (Sherman, 2009), and it is necessary to perform statistical analyses that adjust for measurement error (Greenwald et al., 2015; Greenwald et al., 2009; Kurdi et al., 2018). A failure to do so may be at least partially responsible for weak correlations that have been reported in the past between implicit racial attitudes and behavior. Nevertheless, the IAT has been used successfully to illuminate a wide variety of implicit preferences, including system-justifying preferences concerning race, ethnicity, and social class (Ashburn-Nardo, Knowles, & Monteith, 2003; Hoffarth & Jost, 2017; Horwitz & Dovidio, 2015; Jost et al., 2004;

**Table 1.** Summary of Evidence That Implicit Attitudes Predict Self-Reported Voting Behavior

Study	Political context	Major findings
Bos, Sheets, & Boomgaarden (2018)	Willingness to vote for a populist, radical-right party in The Netherlands, 2013	In a sample of 746 Dutch adults, an ST-IAT (evaluations of a specific populist, radical-right party) predicted self-reported propensity to vote for the party, even after adjusting for demographic, ideological, and other factors.
Lundberg & Payne (2014)	Presidential election in the United States, 2008	In a U.S. sample of 2,013 high- and low-confidence voters, candidate-based AMPs (implicit negative affect associated with the faces of presidential candidates) predicted self-reported voting for Obama versus McCain several weeks later (especially for low-confidence voters), even after adjusting for ideology, partisanship, explicit prejudice, and demographic factors.
Raccuia (2016, Study 1)	Two national referenda (on minimum wage and purchase of fighter jets) in Switzerland, 2014	In a sample of 268 decided and undecided students and employees of the University of Zurich, ST-IATs (evaluations of two proposed national referenda) predicted self-reported voting intentions and behavior, but not always after adjusting for ideology—which was a strong predictor of behavior.
Raccuia (2016, Study 2)	Single-payer health insurance initiative proposed by the Social Democratic Party in Switzerland, 2014	In a sample of 351 decided and undecided students and employees of the University of Zurich, an ST-IAT (evaluation of a public-health-insurance initiative) predicted self-reported voting intentions and behavior, but not always after adjusting for ideology—which was a strong predictor of behavior.
Raccuia (2016, Study 3)	Referendum to limit annual migration sponsored by the Ecology and Population (“Ecopop”) Organization in Switzerland, 2014	In a sample of 457 decided and 82 undecided Swiss adults, an ST-IAT (evaluation of politicians, parties, and organizations associated with “Ecopop”) predicted self-reported voting intentions and behavior, but not always after adjusting for explicit attitudes about immigration—which were strong predictors of behavior.
Ryan (2017)	Presidential election in the United States, 2008	In a sample of 579 “indifferent” and 259 “ambivalent” (but not “one-sided”) U.S. voters, a candidate-based AMP (difference in implicit negative affect associated with the faces of presidential candidates) predicted self-reported voting for Obama several weeks (or months) later, even after adjusting for partisanship and demographic factors.

Note: IAT = implicit association test; ST-IAT = single-target implicit association test; AMP = affective misattribution procedure.

Newheiser, Dunham, Merrill, Hoosain, & Olson, 2014). With respect to organizational behavior, Jost et al. (2009a) recounted 10 studies that business executives would be foolish to ignore, insofar as they demonstrate that implicit attitudes can predict discriminatory decision making. Some, if not all, of those studies should still be taken seriously by anyone who values a workplace that is free of discrimination. Hundreds more studies have now been reviewed quantitatively by Kurdi et al. (2018), as noted above. As research practices improve in terms of scoring algorithms, sample sizes, and measurement correspondence, the case for implicit attitudes is strengthening. Although implicit attitudes do not always explain significant amounts of incremental variance after explicit attitudes are taken into account, they very often do (Kurdi et al., 2018).

Research on implicit attitudes has been especially fruitful for understanding political psychological phenomena. Gawronski, Galdi, and Arcuri (2015) described the results of 20 studies in which implicit attitudes were used successfully to predict subsequent political

judgments and behaviors (see pp. 5, 8, 9). In Table 1, I summarize an additional set of studies demonstrating that implicit attitudes predict self-reported voting behavior and that have appeared since Gawronski and colleagues’ review article went to press. I would also like to draw attention to a groundbreaking study of more than 1,000 Americans, which revealed—in addition to specific candidate, party, and policy preferences—that trust in government can also be measured implicitly and that (even after adjusting for explicit trust in government) the implicit attitude predicts general system-justification scores and national loyalty in anticipated circumstances of natural disaster and foreign attack (Intawan & Nicholson, 2018).

Payne et al. (2017) proposed using aggregate-level indicators of implicit bias at the organizational (rather than individual) level of analysis as a way of assessing whether there may be potential problems in a given organization. Something similar could be attempted at the level of societies and polities to measure the “social climate” (e.g., Charlesworth & Banaji, 2018). A time-series

approach to aggregate measurement would enable researchers to monitor the success of various system-level interventions; that is, to gauge the extent to which “culture-wide changes in implicit prejudice” have (or have not) taken place (Hardin & Banaji, 2013, p. 21).

## Concluding Remarks

A context-sensitive measure of social and political attitudes—including attitudes toward the self, social groups, and the social system—is, at the end of the day, far more useful to social and political psychologists than something resembling an IQ test or a clinical diagnosis would be. We know that specific sociocultural environments condition some evaluative responses much more strongly than others (Banaji, 2001, 2004; Banaji & Greenwald, 2013; Hardin & Banaji, 2013; Ross & Nisbett, 2011), and there is considerable variability across domains when it comes to relations among implicit, explicit, and behavioral responses (Greenwald et al., 2009; Kurdi et al., 2018; Nosek, 2005). Our job is to put all of this together: to understand how dynamic personal factors—such as beliefs, opinions, and values, which are affected by memories and experiences—interact with dynamic social, cultural, and historical factors—such as situational and institutional forces, including new waves of immigration and new forms of legislation—to produce attitudes and behaviors in potentially value-laden contexts. This corresponds exceedingly well to Kurt Lewin’s conception of the “person in the situation,” on which the science of social psychology continues to rest.

## Recommended Reading

- Gawronski, B., Galdi, S., & Arcuri, L. (2015). (See References). A helpful discussion of how implicit measures can be useful for investigating distal sources of political preferences by attenuating self-presentational bias and a comprehensive review of research linking implicit attitudes to subsequent voting decisions.
- Glaser, J., & Finn, C. (2013). How and why implicit attitudes should affect voting. *PS: Political Science & Politics*, *46*, 537–544. A thoughtful analysis of psychological mechanisms that are believed to affect voting and other behaviors and an accessible summary of research showing that both implicit and explicit preferences shape behavior.
- Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009a). (See References). An overview of the historical context in which the measurement of implicit bias arose and a detailed, point-by-point refutation of Tetlock and Mitchell’s various critiques of research on the implicit association test.
- Payne, K., Niemi, L., & Doris, J. M. (2018, March 27). How to think about “implicit bias.” *Scientific American*. Retrieved from <https://www.scientificamerican.com/article/how-to->

think-about-implicit-bias/. A brief, timely, and accessible statement of why research on implicit bias should be taken seriously, despite fallacious interpretations that have become popular by stoking controversy.

## Action Editor

Randall W. Engle served as action editor for this article.

## ORCID iD

John T. Jost  <https://orcid.org/0000-0002-2844-4645>

## Acknowledgments

This article is based on a presentation to the National Science Foundation in November of 2017. I thank Mahzarin Banaji, Kao-Wei Chua, Chris Draheim, Randall Engle, Russell Fazio, Tony Greenwald, György Hunyady, Benedek Kurdi, Brian Nosek, and several anonymous reviewers for providing extremely helpful feedback on earlier drafts of this article.

## Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

## Notes

1. On YouTube there are several recent demonstrations and replications of those classic studies, which show that in many countries today, Black children often prefer White dolls (but not vice versa): <https://www.youtube.com/watch?v=PZryE2bqwdk>, <https://www.youtube.com/watch?v=i20d11fGz-0>, [https://www.youtube.com/watch?v=eFCo87zeb\\_w](https://www.youtube.com/watch?v=eFCo87zeb_w), and <https://www.youtube.com/watch?v=QRZPw-9sJtQ>.
2. Skeptics may claim that when a Black person shows a pro-White, anti-Black bias on the IAT, it is because of a statistical artifact or methodological noise—just a bad measure giving a bad reading. But then it is incumbent on them to explain why the magnitude of out-group favoritism on the part of disadvantaged groups is correlated with measures of system justification, including opposition to equality and political conservatism (Ashburn-Nardo et al., 2003; Hoffarth & Jost, 2017; Jost et al., 2004).

## References

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- 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. doi:10.1207/s15327965pli1504\_01
- Asch, S. (1987). *Social psychology*. Oxford, England: Oxford University Press. (Original work published 1952)
- Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans’ implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*, 62–87. doi:10.1521/soco.21.1.61.21192



- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.
- Banaji, M. R. (2004). The opposite of a great truth is also true: Homage to koan #7. In J. T. Jost, M. R. Banaji, & D. A. Prentice (Eds.), *Perspectivism in social psychology: The yin and yang of scientific progress* (pp. 127–140). Washington, DC: American Psychological Association.
- Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. New York, NY: Delacorte Press.
- Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*, 668–688. doi:10.3758/s13428-013-0410-6
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22. doi:10.1037/0022-3514.87.1.5
- Bartlett, T. (2017, January 5). Can we really measure implicit bias? Maybe not. *The Chronicle of Higher Education*. Retrieved from <https://www.chronicle.com/article/Can-We-Really-Measure-Implicit/238807>
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41. doi:10.1037/0003-066X.61.1.27
- 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.
- Bos, L., Sheets, P., & Boomgaarden, H. G. (2018). The role of implicit attitudes in populist radical-right support. *Political Psychology*, *39*, 69–87. doi:10.1111/pops.12401
- Brody, S., Veit, R., & Rau, H. (1999). Four-year test-retest reliability of self-measured blood pressure. *Archives of Internal Medicine*, *159*, 1007–1008.
- Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, *57*, 278–287.
- Charlesworth, T. E. S., & Banaji, M. R. (2018). *Patterns of implicit and explicit attitudes I. Long-term change and stability from 2007–2016*. Manuscript submitted for publication.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814. doi:10.1037/0022-3514.81.5.800
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2001). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316–328.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637.
- Forscher, P., Lai, C. K., Axt, J., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2016). A meta-analysis of change in implicit bias. *Open Science Framework*. Retrieved from <https://osf.io/b5m97/>
- French, D. (2017, January 10). Implicit bias gets an explicit debunking. *National Review*. Retrieved from <https://www.nationalreview.com/2017/01/implicit-bias-debunked-study-disputes-effects-unconscious-prejudice/>
- Gawronski, B., Galdi, S., & Arcuri, L. (2015). What can political psychology learn from implicit measures? Empirical evidence and new directions. *Advances in Political Psychology*, *36*(1), 1–17. doi:10.1111/pops.12094
- Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*, 300–312. doi:10.1177/0146167216684131
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561. doi:10.1037/pspa0000016
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*, 669–684.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41. doi:10.1037/a0015575
- Hardin, C. D., & Banaji, M. R. (2013). The nature of implicit prejudice: Implications for personal and public policy. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 13–30). Princeton, NJ: Princeton University Press.
- Hoffarth, M., & Jost, J. T. (2017). When ideology contradicts self-interest: Conservative opposition to same-sex marriage among sexual minorities—A commentary on Pinsof and Haselton (2016). *Psychological Science*, *28*, 1521–1524. doi:10.1177/0956797617694866
- Horwitz, S. R., & Dovidio, J. F. (2015). The rich—Love them or hate them? Divergent implicit and explicit attitudes toward the wealthy. *Group Processes & Intergroup Relations*, *20*, 3–31. doi:10.1177/1368430215596075
- Intawan, C., & Nicholson, S. P. (2018). My trust in government is implicit: Automatic trust in government and system support. *Journal of Politics*, *80*, 601–604. doi:10.1086/694785
- 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. doi:10.17605/OSF.IO/6UE35
- Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009a). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should

- ignore. *Research in Organizational Behavior*, 29, 39–69. doi:10.1016/j.riob.2009.10.001
- Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009b). An invitation to Tetlock and Mitchell to conduct empirical research on implicit bias with friends, “adversaries,” or whomever they please. *Research in Organizational Behavior*, 29, 73–75.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology*, 81, 774–788. doi:10.1037//0022-3514.81.5.774
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., . . . Banaji, M. R. (2018). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *PsyArXiv Preprints*. Retrieved from <https://psyarxiv.com/582gh/>
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81, 842–855. doi:10.1037//0022-3514.81.5.842
- Lundberg, K. B., & Payne, B. K. (2014). Decisions among the undecided: Implicit attitudes predict future voting behavior of undecided voters. *PLOS ONE* 9(1), Article e85680. doi:10.1371/journal.pone.0085680
- Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 164–195). Hoboken, NJ: Wiley.
- Newheiser, A. K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, 50, 1081–1090. doi:10.1037/a0035054
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565–584. doi:10.1037/0096-3445.134.4.565
- Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition & Emotion*, 22, 553–594.
- Nosek, B. A., & Sriram, N. (2007). Faulty assumptions: A comment on Blanton, Jaccard, Gonzales, and Christie (2006). *Journal of Experimental Social Psychology*, 43, 393–398.
- Nunes, K. L., Firestone, P., & Baldwin, M. W. (2007). Indirect assessment of cognitions of child sexual abusers with the implicit association test. *Criminal Justice and Behavior*, 34, 454–475.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108, 562–571. doi:10.1037/pspa0000023
- 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.
- 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. doi:10.1016/S0005-7967(03)00018-4
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28, 233–248. doi:10.1080/1047840X.2017.1335568
- Project Implicit. (2011). *Ethical considerations*. Retrieved from <https://implicit.harvard.edu/implicit/ethics.html>
- Raccuia, L. (2016). Single-target implicit association tests (ST-IAT) predict voting behavior of decided and undecided voters in Swiss referendums. *PLOS ONE*, 11(10), Article e0163872. doi:10.1371/journal.pone.0163872
- Richmond, S. D., Rees, G., & Edwards, S. J. L. (Eds.). (2012). *I know what you're thinking: Brain imaging and mental privacy*. New York, NY: Oxford University Press.
- Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. London, England: Pinter & Martin.
- 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, 437–465.
- Ryan, T. J. (2017). How do indifferent voters decide? The political importance of implicit attitudes. *American Journal of Political Science*, 61, 892–907. doi:10.1111/ajps.12307
- Sherman, J. W. (2009). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 391–426). Hillsdale, NJ: Erlbaum.
- Singal, J. (2017). Psychology's favorite tool for measuring racism isn't up to the job. *New York Magazine*. Retrieved from <http://nymag.com/scienceofus/2017/01/psychology-racism-measuring-tool-isnt-up-to-the-job.html>
- 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.