



ANNUAL REVIEWS Further

Click [here](#) for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Unconscious Racism: A Concept in Pursuit of a Measure

Hart Blanton¹ and James Jaccard²

¹Department of Psychology, Texas A&M University, College Station, Texas 77843;
email: hblanton@gmail.com

²Department of Psychology, Florida International University, Miami, Florida 33199;
email: jjaccard@gmail.com

Annu. Rev. Sociol. 2008. 34:277–97

First published online as a Review in Advance on
April 17, 2008

The *Annual Review of Sociology* is online at
soc.annualreviews.org

This article's doi:
10.1146/annurev.soc.33.040406.131632

Copyright © 2008 by Annual Reviews.
All rights reserved

0360-0572/08/0811-0277\$20.00

Key Words

racism, implicit association test, implicit attitudes, psychometrics

Abstract

It is common in scientific and popular discussions to claim that unconscious racism is both prevalent and potent in modern societies. We review the theoretical models that posit different forms of unconscious racism and evaluate the empirical evidence for them. Our analysis suggests that people may sometimes lack knowledge of and control over the causes and consequences of their racial biases. However, there is little evidence to support the more provocative claim: that people possess unconscious racist attitudes. Many of the arguments to the contrary rest on strong interpretations of response patterns on implicit attitude measures. Although advances in implicit measurement can improve our understanding of racial bias, at present their use as tools for rooting out unconscious racism is limited. We describe research programs that might move these constructs to firmer scientific footing, and we urge inferential caution until such research programs are carried out.

Implicit measure: an indirect measure that does not require declarative knowledge of one's standing on the construct being assessed

Implicit association test (IAT): an experimental method designed to measure the strength of associations linking social categories (e.g., blacks versus whites) to descriptive or evaluative anchors (e.g., good versus bad)

INTRODUCTION

There are indications that the landscape of racism in the United States is changing. Many scholars accept that there has been a precipitous decline in old-fashioned racism (also known as dominant racism and redneck racism). Old-fashioned racism is characterized by explicitly racist attitudes regarding the inferiority of blacks or other minorities and overt tendencies to engage in unambiguously discriminatory behavior. Trends in national surveys suggest that such racist leanings are on the decline and that endorsement of racial equality as a societal goal is widely shared. In our seemingly more tolerant society, however, evidence of racism remains easy to find. Racial disparities can be found in a range of outcomes related to employment, income, education, and health. A recent survey also found that nearly half of the black people in the general American population reported an experience with discrimination in the previous week (Gallup Org. 2001). There are various interpretations of the gap between survey trends and racial disparities, one of which is simply that survey respondents are being less truthful about their racist attitudes than they once were.

Here we focus on a more provocative account: that racist attitudes remain prevalent but now are buried in the unconscious. As a result of this migration, racist sentiments can no longer be detected by traditional survey instruments. Interest in the concept of unconscious racism has taken hold in many psychological circles, and it has even captured the attention of the news media and the popular press (e.g., Shermer 2006, Vedantam 2005). Concern about combating unconscious racism has also begun to influence legal debates on the proper function of antidiscrimination policies (Kang & Banaji 2006, Mitchell & Tetlock 2006). When this topic was reviewed recently for an article in this series, Quillian (2006, p. 323) concluded that "an exclusive focus on conscious beliefs in research about prejudice and discrimination misses an important source of discriminatory behavior, because in some cases the perpetrator of discrimination may not

be aware of how their implicit beliefs about race influence their judgments and actions."

We offer a close examination of the concept of unconscious racism and the empirical evidence for it. Our thesis is that current claims of a pervasive and unconscious form of racism rest on overly aggressive interpretations of a new class of psychological inventories, termed implicit attitude measures. We develop our argument by considering the ways in which racism may be unconscious and the evidence for these different forms of unconscious racism. Our review shows that strong inferences regarding the presence and prevalence of unconscious racism are only warranted if one is willing to make strong inferential leaps. We then examine the most popular new implicit measure, the implicit association test (IAT), and discuss the considerable challenges facing this new research domain. For ease of presentation, we focus our attention on the form of unconscious racism that ostensibly is held by many white Americans against black Americans. However, our critique applies to other forms of unconscious racism and can be expanded to include ethnocentrism, sexism, and most other forms of intolerance.

DEFINING UNCONSCIOUS RACISM

In a statement that is representative of many claims found in the contemporary psychological literature, Dovidio & Gaertner (2004, p. 4) suggest that

because of a range of normal cognitive, motivational, and sociocultural processes that promote intergroup biases, most whites also develop some negative feelings toward or beliefs about blacks, of which they are unaware or which they try to dissociate from their non-prejudiced self-images.

But what is the evidence that most people are not aware of their biases? What is the evidence that "the vast majority of white Americans harbor unconscious negative associations about blacks" (Dovidio & Gaertner 2004,

p. 20) or that “any non–African American subgroup of the United States population will reveal high proportions of persons showing statistically noticeable implicit race bias in favor of [European Americans] relative to [African Americans]” (Greenwald & Krieger 2006, p. 945)?

The answer seems to depend on how one defines unconscious. In the following sections, we consider three distinct definitions of unconscious racism using a framework adapted from articles by Gawronski & Bodenhausen (2008), Gawronski et al. (2006), and Olson & Fazio (2004). Two of these definitions point to forms of unconscious bias that can have significant consequences in the lives of individuals but that can also be accommodated within traditional psychological models. The third definition is more provocative and is, we believe, the definition that most lay people and scholars have in mind when they consider the term unconscious. As we show, the first two definitions enjoy some empirical support, but the third is tenuous.

First Definition: Unconscious = Unknown Effects

The first working definition of unconscious refers to individuals’ lack of awareness of the effects of their own actions on other people, social institutions, and so on. When unconscious racism is framed in these terms, it draws attention to the unwitting ways in which one’s own actions promote racial disparities. Such unconscious racism has been described in sociological theories that refer to institutional racism (e.g., Carmichael & Hamilton 1967). Such frameworks show how people become embedded within racist institutions such that they may fail to see how their adherence to accepted social norms inadvertently reinforces existing inequalities. This version of the unwitting (unconscious) racist in some ways resembles the unwitting actor portrayed in the theory of symbolic racism (Sears & Kinder 1971). Symbolic racism theory advances the view that individual tendencies to endorse traditional American values, including individualism and self-reliance,

can cause whites to act in ways that promote racial disparities.

We do not contest the claim that people often fail to see the racist implications of their own actions, and we agree that many types of ignorance need to be studied due to their societal implications. However, we find definitions of unconscious racism that are based on this observation to be unsatisfying. To say that an actor is unconsciously racist (or prejudiced or biased) on such grounds simply draws attention to the fact that ignorance can lead to unintended consequences. This is not news. Moreover, for a scholar to label an action unconscious by reference to this form of ignorance, he or she must necessarily view the action from a position that is unavailable to or disputed by the actor. One way of gaining such a vantage point is to adopt a different value system. Thus, researchers who seek to root out unconscious racism by drawing attention to unintended racist outcomes often must justify why their own value systems (e.g., those that emphasize equality) should take intellectual priority over those of other actors (e.g., those that emphasize individualism). In so doing, these researchers move the discussion outside the realm of science and into the realms of philosophy and politics. Thus, although we doubt that many scholars would dispute the claim that people sometimes fail to appreciate the (racial) consequences of their own actions, attempts to invoke this claim to label a specific actor as unconsciously racist are subject to controversy (see Tetlock 1994).

Second Definition: Unconscious = Unknown Causes

Another way in which people may lack conscious access to their racist leanings is that they may fail to perceive the factors that cause them to exhibit racial preferences. People are unconscious racists in this sense if they are unable to gain subjective access to the determinants of their own actions. A large literature on impression formation offers empirical support for this perspective. Psychologists have shown, for instance, that a shove given by one person to

Subliminal priming:
the process of
activating a stimulus or
concept outside of
conscious awareness

another might appear aggressive when the person doing the shoving is black, but that it might seem playful when the person doing the shoving is white (Duncan 1976). Had the shove been unambiguous, however—such that it was clearly aggressive or clearly playful—race would not exert such an effect. The role of ambiguity in the person perception literature is important because it suggests that people might be influenced by a person's race but might fail to recognize this influence. Researchers thus have suggested that social stereotypes fill gaps in meaning when the implications of an action or event are unclear, thereby causing biased reactions to seem objective, rational, and justified (Kunda & Thagard 1996).

Additional evidence of this effect is found in studies that activate stereotypes outside of conscious awareness. In an influential investigation, Devine (1989) used a computer-based task to subliminally prime participants with words related to black stereotypes. These words were presented for 80 milliseconds and were then masked by a jumbled series of letters. After exposure, participants read a paragraph describing an actor who engaged in an ambiguously hostile action. The race of this individual was not stated, but after being primed with black racial words, participants were more likely to interpret the ambiguous behavior as hostile.

Devine's research and the many other subliminal priming studies inspired by her work are often invoked as evidence that people have an unconscious tendency to hold racially biased perceptions. According to this logic, if researchers can bias social perceptions simply by exposing participants to racial cues (over a time period measured in milliseconds), then it might be common for individuals to be heavily influenced by racial stereotypes encountered in everyday life. Consider, for instance, how the mere act of interacting with a black person could activate racial stereotypes among whites, which could thereby cause them to perceive more negative qualities in black individuals than in white individuals.

However, one should not overinterpret Devine's findings. Although people certainly

are not conscious of all the factors that influence their perceptions and actions, this fact does not necessarily indicate that there is an epidemic of untapped and largely unconscious racism in American society. In her study, Devine used priming words that activated unusually negative racial stereotypes, stereotypes that would be rejected by many if they were to encounter and process them consciously. (Her studies used such value-laden words as "Harlem," "prejudice," "ghetto," "welfare," "unemployed," and "nigger.") Thus, it is questionable to suggest that Devine tapped into the "normal" unconscious racist tendencies of the individuals she studied. In fact, research participants respond differently if they are primed with words that are less racist in nature, such as "black," "ethnic," "afro." A set of priming studies that used these exemplars activated responses that were consistent with individuals' consciously endorsed attitudes: Only individuals who had reported higher levels of racial prejudice prior to priming made more negative evaluations of target individuals following race priming (Lepore & Brown 1997). This finding suggests that even when people are not conscious of the factors that cause them to act, they may nonetheless act in ways that largely reflect their consciously held attitudes.

Reconsider the study by Duncan (1976) in light of this analysis. Duncan's investigation showed how an ambiguous shove could appear aggressive when it was associated with a black actor as opposed to a white actor. In all likelihood, the participants in Duncan's study knew that social stereotypes linked blacks to violence. It also is possible that some of the participants in this study consciously endorsed this view, although some may have done so with a degree of internal conflict. If such was the case, it would be incorrect to say that these participants were not "conscious" of the racial stereotypes or beliefs that ultimately influenced their perceptions. Nevertheless, study participants might have believed (incorrectly) that they were not being influenced by these views. Because knowledge about the presence of specific attitudes and beliefs does not always

translate into knowledge about the effects that these same attitudes and beliefs exert on perceptions and behaviors, a substantial literature has arisen that seeks to determine the conditions under which people can and cannot correct for their known discriminatory tendencies (see Wegener & Petty 1997).

Another vantage point for viewing the influence of subliminal primes can be found in Bargh & Pietromonaco (1982). These researchers subliminally primed participants with aggressive words (e.g., “hostile,” “rude”), causing the participants to draw more aggressive interpretations about an actor who was engaged in an ambiguously aggressive action. The effects were similar to those reported by Devine (1989), but we doubt many would find utility in conceptualizing this as a study of “unconscious aggressive attitudes.” Bargh & Pietromonaco activated a shared social concept—aggression—and in so doing they influenced the momentary perceptions of their participants. Similarly, a researcher who primes racial bias might exert momentary influences on a group of individuals, but this effect could simply show that human beings are open to (unconscious) suggestion—not that they hold any particular hidden biases that are waiting to spring forth.

In short, although it appears that people can be clueless about the factors that influence their actions, such observations are not unique to the study of racism, nor do they create a firm basis for claiming that most people have unconscious racist leanings. Indeed, if one were to pursue this type of justification for labeling an event unconscious, then a large number of research domains within experimental psychology would need to be redefined as the study of the unconscious. Experimental psychologists typically eschew introspection as a research method and instead rely on methods that subtly control the invisible forces that lead to interesting effects. They might, for instance, alter the framing of a social dilemma to study the cognitive heuristics that influence decisions (e.g., Kahneman & Tversky 1973), or they might manipulate a person’s behavior to trigger a specific motivation (Festinger 1957). In studies such as

these, researchers often rely on deception or use between-subject designs to ensure that participants remain in the dark about the factors thought to influence outcomes. The same research methods could also be used to study egalitarian tendencies, if this were in the interest of a researcher (e.g., priming concepts related to racial harmony to study their effects on interracial cooperation). Thus, although we agree that there are times when people are unable to identify the factors that lead them to act in racially biased ways, this observation should not be overinterpreted. Similar claims could be made for many other perceptions and behaviors.

Third Definition: Unconscious = Inaccessible Attitudes

Although it is feasible to argue that people lack knowledge of both the causes and the consequences of their actions, this set of observations alone does not provide a basis for claiming that people possess racist attitudes that escape subjective awareness. What is the evidence for this third type of unconscious racism—ignorance about the presence of racist attitudes? The study of this phenomenon raises a particularly vexing problem. How do researchers show that people possess views that they cannot perceive?

Rationalized racism. One strategy has been to show that people seem to hide their true beliefs from themselves and from others through rationalization (Dovidio & Gaertner 2004). A study by Gaertner & Dovidio (1977) is illustrative. These researchers created a laboratory situation in which white participants encountered either a white or a black confederate in need of help. When participants were alone and thus were the only ones who were in a position to help, the majority (over 85%) helped. In this case, the race of the distressed individual exerted no influence on helping rates. However, when the distressed individual was in proximity to other nonresponsive bystanders, a race effect emerged. A majority still helped the white

confederate (75%), but only a minority helped the black confederate (38%). Gaertner & Dovidio interpreted this as evidence that whites will act in a racist manner if they can maintain the belief that their actions were due to some other cause (e.g., "no one else was helping and so I did not think it was an emergency"). Numerous studies have documented similar tendencies by whites to act in a racially biased manner against blacks, but only when their actions can be attributed to nonracist concerns (see Aberson & Ettlin 2004).

Although studies suggesting rationalization may help explain how racism manifests itself in modern societies, such findings do not necessarily support the conclusion that people lack conscious access to their racist evaluations. In fact, attempts to make such arguments are logically problematic: How could people try to disguise their racist leanings if they did not possess some insight into these same racist leanings? For researchers to assert the presence of an attitude that is hidden from the very person who holds it, it seems that they must find a way of indexing attitudes that (*a*) are not cognitively accessible to the individual and that (*b*) operate outside of the individual's awareness, but that (*c*) nonetheless exert consequential influences on the individual's behavior in the form of racial discrimination. This is a tall order.

Implicit racism. To meet this challenge, researchers have drawn on cognitive research studying implicit memory. This research shows that people can develop knowledge of procedures and a familiarity with stimuli without also developing explicit memory of the events that produced these effects (e.g., Schacter 1987). The term implicit attitudes refers to attitudes that cannot be directly perceived and must therefore be measured indirectly (Banaji 2001, Greenwald et al. 1998). This contrasts with (traditional) explicit attitudes that exist in declarative memory and that thus can be assessed via self-reporting. Some researchers contend that a person might possess one set of conscious attitudes that support racial tolerance and another set of unconscious attitudes that are racially bi-

ased (Banaji 2001, Greenwald et al. 1998). The gap between conscious attitudes (measured explicitly) and unconscious attitudes (measured implicitly) is commonly invoked as an explanation for the survey trends noted at the beginning of this article. Although racism appears to be on the decline when assessed with explicit measures, it seems to remain strong and prevalent when assessed with implicit measures (see Dovidio & Gaertner 2004, Nosek et al. 2007).

The devil is in the details. Social psychology's embrace of cognitive models of memory (Schacter 1987) and learning (e.g., Zajonc 1968) for the purpose of measuring implicit racism reflects an interesting intellectual turn. Although psychologists' interest in the unconscious can be traced to Freud (1901), whose thinking was incorporated into early models of unconscious racism (e.g., Kovel 1977), Freud's hold over academic psychology diminished quickly because the constructs he invoked were subject to weak measurement and encouraged post hoc explanations. With the rise of cognitive models that drew on the concepts of implicit memory and implicit learning, and with access to more sophisticated computer-based measurement strategies, researchers are trying again to mine the unconscious for explanations of human behavior. As with many research fads, however, the current enthusiasm for this approach is giving way to a greater appreciation for the complexities involved. Perhaps the greatest challenge facing this discipline is to form the basis by which one might assert that responses measured with an implicit inventory reveal the workings of an unconscious attitude.

It thus is useful to consider why explicit measures should be unable to tap the unconscious. Explicit attitude measures, by design, invoke two general processes when they are used to assess an attitude: (*a*) They ask respondents to reflect consciously on their own attitudes and (*b*) they ask respondents to report their conclusions from this moment of self-reflection. In contrast, implicit measures try to bypass the conscious mind: (*a*) They unobtrusively activate an attitude so that it need not be consciously

perceived and (b) they unobtrusively assess this evaluation so that it cannot be consciously obfuscated. Although dozens of methods for implicitly assessing attitudes have been suggested in the last 10 years (see Olson & Fazio 2004), two common types of procedures make up the bulk of such efforts. One of these relies on priming. With priming, participants engage in a task that is designed to prime an attitude outside of conscious awareness. This evaluation is then assessed unobtrusively, typically using a response-latency task. The second common strategy relies on mental categorizations. For example, the IAT (described in detail below) assesses the time it takes for people to classify stimuli into different evaluative categories, with the response-latency tasks crafted in such a way that they ostensibly index the person's underlying attitude.

Consider as an example a priming-based method developed by Fazio et al. (1995), which is often referred to as affective priming. Although some have questioned the robustness of the underlying effect (Spruyt et al. 2002), studies indicate that people are quicker to identify a word if it has been preceded by a prime of the same valence, as opposed to a different valence. Fazio et al. (1995) thus developed an implicit attitude measure that uses subliminal race primes to assess the degree to which the primes facilitate recognition of positively valenced words (which implies that the primed race is positively evaluated) or negatively valenced words (which implies that the primed race is negatively evaluated). Although it is reasonable to infer that differences in such responses are in part reflective of a person's underlying attitudes, it is an unnecessary leap to assume that the attitude being assessed lives only in the unconscious (a point acknowledged by Olson & Fazio 2004).

Moreover, an affective priming measure sorts people along a psychological dimension by harnessing an effect that we reviewed above: the tendency to be influenced by an event outside of conscious awareness. We have already noted that subliminal primes might activate consciously held beliefs or external social conceptions. Similarly, implicit measures might unobtrusively activate attitudes that can

be consciously perceived (see Nier 2005). Alternatively, they might sort people by their level of exposure to events that reinforce external social conceptions (see Karpinski & Hilton 2001). It thus seems problematic to infer that a measure assesses an unconscious attitude simply because the assessment technique triggers responses outside of conscious awareness. More research is required to demonstrate the validity of this technique as a method of assessing attitudes that exist outside of conscious awareness.

Perhaps an alternative way to assess the unconscious would be to develop implicit measures that tap responses that people cannot control, making it impossible for them to obfuscate their true (hidden) evaluations. Although popular interest in unconscious racism probably is driven in large part by its ability to challenge popular notions of self-determination and rational choice (see Bargh 2004), it is problematic to conflate the concept of control with the concept of consciousness. Consider the many maladaptive behaviors that result from people's inability to control their impulses. These include drug addiction, obesity, sexual risk taking, and problem gambling. People struggling with these behaviors probably have conscious access to many of the causes of their problems and know the content of their unwanted impulses. However, such knowledge does not translate into control over behavior. Thus, although personal control represents an important area of research in the study of racial bias (see Payne 2001, Correll et al. 2007), lack of control does not provide a viable line for delineating between consciously and unconsciously held racist attitudes. Lack of control does not equate with unconsciousness, although many implicit measures are structured to tap responses that people have difficulty controlling.

Defensible (but still problematic) alternative approaches. Despite these problems, implicit measures are the only hope of gaining access to unconscious attitudes: People cannot report on their unconscious attitudes, so implicit cleverness is needed to gain traction on these hypothesized constructs. But how would

a researcher know if a given implicit response is indexing a mostly unconscious attitude or some other unconscious construct? We suggest two ways of gaining perspectives on this question. Both have been given some attention but are as yet underdeveloped.

The first approach is to show that implicit measures predict discrimination-related criteria, after conscious attitude content has been controlled. The logic of this strategy is that any unique variance claimed by implicit measures might reflect attitudes or beliefs that are not accessible to conscious reflection. However, this approach has its difficulties. To pursue the logic of this approach, a researcher must include a battery of explicit measures that are valid, reliable, and exhaustive. To date, few criterion-prediction studies control for self-reported attitudes, and, when they do, they rely on measures and conceptions of explicit attitudes that were long ago rejected by contemporary attitude theorists as not being viable (e.g., Ajzen 2005, Fishbein & Ajzen 1975; also see Jaccard & Blanton 2007, Payne et al. 2008). Such research effectively sets up a straw man by using inadequate explicit attitude theories that few investigators would embrace. Studies using implicit attitude measures should also control for a wide a range of other constructs that influence behavior and that participants can consciously make note of (e.g., social norms, perceived barriers to action). Even after exhaustive measurement of such constructs, researchers must still make a leap of faith when they claim the presence of an unconscious attitude. That is, they must claim that all relevant conscious processes were measured adequately and were statistically controlled. This is a bold claim.

The second approach is to link implicit measures to neurological correlates that could inform us of the possibility of an attitude or cognition having a cognitive representation. For example, research on cognitive neuroimaging has begun to isolate the neurological correlates of unconscious and conscious visual stimuli (Dehaene et al. 2001), and the use of such methods to decode implicit measures could yield greater theoretical clarity. By neuroimaging dif-

ferent areas of the brain, it may be possible to identify stimuli that have been processed by the individual, in such a manner that they could not possibly be perceived consciously. Such studies are emerging in the literature. For instance, Phelps et al. (2000) linked implicitly measured attitudes (and not explicit measured attitudes) to greater amygdala activation during exposure to black as opposed to white faces. (Amygdala response is associated with processing of emotion; see also Richeson & Shelton 2003.) However, this research is in its infancy and considerable work is needed before researchers can say with confidence that responses to a given implicit measure reflect the presence of an evaluation that is cognitively represented but not consciously accessible.

PUSHING FORWARD

The previous analysis of research on unconscious forms of racism suggests that researchers should exercise caution before suggesting that a given individual or group of individuals holds unconscious racist attitudes. Although people are probably not conscious of some of the causes and consequences of race-related perceptions and actions, the claim that people hold unconsciously racist attitudes is tenuous. We do not make this point to dampen enthusiasm for developing implicit measures, but we think that guarded interpretations and high empirical standards are called for. Such has not always been the case: Some (although certainly not most) of the more influential researchers in this area have made strong claims about the merits of their implicit measures. Therefore, we take a closer look at implicit attitude measurement to reveal some of the complexities and challenges facing this new measurement movement. It is beyond the scope of this review to detail the dozens of implicit attitude measures recently advanced to measure racial biases. Instead, we use one measure as a case study for describing the bold theorizing surrounding it, as well as issues one must consider to develop valid and viable implicit attitude measures. The measure we examine is the IAT—by far the

most popular and empirically assessed of the new instruments. For ease of presentation, we focus on a version that was designed to assess (unconscious) negative evaluations of black people relative to white people, the race IAT (Greenwald et al. 1998).

Implicit Association Test Methodology

In the race IAT, respondents are shown stimuli on a computer screen, which they place into different categories by pressing a key on the right or left side of the computer keyboard. Half of the stimuli pertain to racial groups (white versus black) and half pertain to evaluative words or attributes. The racial stimuli are typically pictures of white or black faces or stereotypic African American or European American names (e.g., Tyrell versus Chip). The evaluative stimuli are words that are positive in character (e.g., “freedom,” “love,” “peace”) or negative in character (e.g., “abuse,” “filth,” “murder”). After a set of practice trials, the IAT presents two types of judgment tasks, and the latencies it takes individuals to make judgments are recorded across 60 trials of each type of task.

One task, often termed the compatible task, is designed to be easy for a person who harbors negative (racist) implicit attitudes about black people relative to white people. Respondents are shown stimuli from the different categories and are instructed to press one key to indicate if the stimulus shown is a photo/name of a white person or a positive word, and another key if the stimulus is a photo/name of a black person or a negative word. The other task, often termed the incompatible task, is designed to be hard for a person who harbors negative (racist) implicit attitudes toward black people relative to white people. Respondents press one key to indicate if the stimulus presented is a photo/name of a black person or a positive word, and another key to indicate if the stimulus is a photo/name of a white person or a negative word. The logic underlying the two tasks is that if respondents hold implicit biases against black people relative to white people, they should be slower to respond

to the incompatible task than the compatible task (i.e., participants should find it harder to perform a task that is incompatible with negative racial stereotypes than a task that is compatible). Researchers then subtract the average response latency for the compatible task from the average response latency for the incompatible task to yield an IAT score. (We discuss qualifications below.) Positive scores are interpreted as evidence of an implicit preference for whites relative to blacks and negative scores are interpreted as evidence of an implicit preference for blacks relative to whites.

Strong Claims

The race IAT has received an unusual degree of popular attention, so much that at times it is difficult to separate the impact it has had due to empirical scrutiny from the impact it has had due to media interest. The measure has been featured on news programs on all the major television networks and in news articles circulated by the Associated Press, and it has been incorporated into sensitivity training workshops sponsored by corporations, community groups, colleges, and even some government agencies. In the popular press, the architects of the IAT have promoted the view that it can be used to tap the unconscious roots of racism that explicit (self-report) attitude measures cannot assess (e.g., Shermer 2006, Vedantam 2005). Similar views of this measure were promoted when the IAT was featured in the popular press book *Blink* (Gladwell 2005).

Much of the interest in this measure also generates from a demonstration website (<https://implicit.harvard.edu>) funded by the National Science Foundation and maintained by Harvard University. This website offers different versions of the IAT to the public and provides respondents with psychological feedback about their implicit biases. Since its inception, the website has provided feedback to the lay public over five million times (Nosek et al. 2007). According to statistics posted on the website, the vast majority of people taking the race IAT (70%) are told that they have an

“automatic preference for Whites over Blacks” and 27% are told that they have a “strong automatic preference for Whites over Blacks.” These diagnoses probably lead many individuals to infer that they possess hidden antiblack racist attitudes.

This apparent epidemic of implicitly racist sentiments against blacks has been taken as evidence of a great social ill that must be remedied. Legal scholars have suggested that the forms of unconscious bias measured by implicit measures reveal social problems that current antidiscrimination laws are not well equipped to solve (Bagenstos 2007), with some researchers suggesting that public policies must be implemented to promote and ensure social equality until and unless evidence of bias on the race IAT and other implicit measures diminishes (e.g., Kang 2005, Kang & Banaji 2006). One scholar (Saujani 2003) has gone so far as to suggest that the IAT could help reduce the level of racism in governing bodies if it was used as a tool to uncover “legislator’s [sic] reliance on unconscious racial stereotypes,” and another (Ayres 2001, pp. 424–35) has argued that IAT scores could “be used as a criterion for hiring both governmental and nongovernmental actors,” ostensibly as a means of counteracting discrimination in the workplace.

Weak Evidence

Strong claims require strong evidence. Perhaps because the claims described above are so provocative, disputes have erupted in both psychological and legal journals, with the proper interpretation of IAT data receiving the most attention. Using the IAT as an illustrative study, we now turn to the special challenges confronting implicit measures and use the IAT as an illustrative example.

Test validity: criterion prediction. In a large meta-analysis of the IAT as a predictor of criteria, Greenwald et al. (2008, p. 2) concluded that the “IAT measures significantly exceeded self-report measures in predictive validity.” Despite such conclusions, a careful

review of this literature yields a mixed picture. First, the measurement and control of explicit attitudes in these studies are universally weak, such that any claims that implicit measures predict criteria over and above explicit measures are dubious. The effects of the IAT on discrimination-related criteria also seem to be moderated by many factors. For example, Perugini et al. (2007) found that higher IAT scores predicted more guilty judgments of a black (Afro Caribbean) defendant in a hypothetical case study, but they found this linear relationship only if (a) attention was drawn to the self (by first having participants answer questions that focused their attention on their unique qualities), and participants also scored high on a measure assessing their private concern for acting prejudice (e.g., “I get angry with myself when I have a thought or feeling that might be considered prejudiced”) or if (b) attention was not drawn to the self and concern for acting prejudice was low. Similarly, Ziegert & Hanges (2005) found that the IAT predicted more negative evaluations of black job applicants, but only if research participants were explicitly instructed to discriminate against blacks.

Perhaps as a result of the many moderating factors, the effects of the IAT vary considerably across research studies. For instance, Shelton et al. (2005) found that higher antiblack IAT scores among white participants predicted more positive interactions with black students (as rated by the black students themselves), whereas McConnell & Leibold (2001) found that these same scores predicted more negative interactions by white participants interacting with a black experimenter (as rated by the black experimenters and independent raters). But Vanman et al. (2004) found that white participants’ race IAT scores did not predict their ratings of black applicants in a simulated graduate admissions study, even though another implicit measure they examined (which assessed activation of the facial muscles associated with positive and negative affect) was predictive of these ratings. To the extent that a single trend can be discerned in the broader

literature using implicit measures, that trend is for some measures to predict subtle forms of bias (e.g., eye contact, facial expressions) in laboratory settings rather than direct forms (e.g., spoken comments; see Dovidio & Gaertner 2004). Even here, however, there is debate about the meaning of these effects (Olson & Fazio 2008).

The criterion-prediction studies also have methodological limitations that caution against vigorous applied claims or suggestions that most people possess consequential implicit biases. For example, in the most frequently cited study of behavioral predictability of the race IAT (McConnell & Leibold 2001), the data were analyzed so as to mask an aggregate-level disconnect between implicit attitudes and behavior (Blanton et al. 2008a): Whereas 90% of the sample had IAT scores that implied some form of racial bias, 70% of the sample behaved more favorably toward a target black person as compared to a target white person. Also, the modest correlation between implicit attitudes and behavior in this laboratory study was outlier driven (by the inclusion of a single middle-aged woman with unusually slow IAT scores relative to the other college students in the sample). Despite such limitations, this area remains an important avenue for research on the determinants of racial biases.

Test validity: known confounds. Systematic error variance poses a concern for any psychometric inventory, but the IAT faces a wide range of new validity challenges that are representative of those facing other implicit attitude measures. Although self-reports can be threatened by social desirability biases and a wide range of response artifacts, decades of research have provided researchers with methods for minimizing and addressing these concerns. The validity threats that challenge the IAT are unfamiliar to researchers by comparison, and so they will require greater attention in the coming years. Here we review the more common threats.

General processing speed. One threat applies to all implicit measures that rely on response la-

tencies: the potentially confounding influence of a construct we have termed general processing speed (GPS). Irrespective of their attitudes (implicit or otherwise), some individuals respond faster than others on a wide range of cognitive tasks due to the faster “processing speed” of their brains. There is a substantial literature in cognitive psychology on GPS, and studies have found it to be correlated with (or confounded with) such individual difference variables as intelligence, age, alcohol use, a host of cognitive deficit variables, and even attitude constructs like explicit racism (see Blanton et al. 2006, 2008b). Because GPS affects response latencies, any measure that infers attitudes from reaction time data must contend with such confounds. With the IAT, for instance, there is ample evidence to indicate that GPS affects both the compatible and incompatible tasks such that faster processing speed is associated with quicker responses to each task. In our studies, we find that GPS typically accounts for 25%–50% of the variance in responses on a given task. Typically, GPS also is correlated with the IAT difference score when the mean compatible response latency is subtracted from the mean incompatible response latency.

The influences of GPS in latency-based measures have been downplayed, despite the extensive attention GPS has received in cognitive psychology and despite its clear psychometric implications. A one-second difference in average latencies on the compatible and incompatible tasks of the IAT implies different amounts of implicit prejudice for someone who generally responds slowly to stimuli than for someone who generally responds quickly to stimuli. A viable psychometric model of implicit attitudes that relies on latencies must take this interaction between GPS and implicit bias into account, yet none of the current measures does so in any formal way. The architects of the IAT have expressed concern for GPS, but only as a general confound. Indeed, their attempts to deal with it (Greenwald et al. 2003) may have created more problems than have been solved, as we discuss here.

Almost all implicit attitude measures employ multiple test trials. For example, in the IAT, individuals respond to 60 compatible trials and 60 incompatible trials, with the individually calculated mean latency of each of the 60 trial types (compatible versus incompatible) serving as the basis for an IAT score for that individual. We refer to these scores as block means. It is well known that latencies on trials within a block can vary because of random “noise.” Trial error refers to the nonsystematic error that causes an individual to respond faster or slower to different trials within a given block, independent of prejudice. Trial error might occur because a particular stimulus on a given trial is unusually attention-grabbing or because of momentary distractions (e.g., loud noises, distracting thoughts) in the testing environment. Researchers seek to control and minimize such random noise. By averaging latencies across trials, the effects of trial error are assumed to cancel. Thus, the trials act like multiple items on a traditional inventory that, when averaged, produce a more reliable estimate of the underlying construct.

Partly in order to take GPS into account, Greenwald et al. (2003) advocated scoring the IAT by dividing the difference between the mean incompatible and compatible blocks by an individually derived standard deviation, namely

$$\text{IAT} = (\text{IRL} - \text{CRL}) / \text{SD}_{\text{WI}}, \quad 1.$$

where IRL is the mean incompatible response latency across the 60 trials for that individual, CRL is the mean compatible response latency across the 60 trials for that individual, and SD_{WI} is a within-individual standard deviation calculated across the compatible and incompatible trials. This is the scoring algorithm that now enjoys widespread use.

The part of the IAT score that accommodates GPS must lie in SD_{WI} . However, this term is only modestly related to processing speed, if at all (Blanton et al. 2007, 2008b). Further, closer examination of Equation 1 reveals other psychometric peculiarities. SD_{WI} is calculated

within one individual across both the compatible and the incompatible trials. As a result, if a researcher is successful in eliminating random noise across trials, then each of the incompatible trials will yield the same latency score and all will equal the block mean value (IRL). The same dynamic would lead to CRL. It can be shown mathematically that in the absence of random noise, SD_{WI} must equal half the difference of the two block means ($\text{IRL} - \text{CRL}$). This means that as random noise is minimized, IAT scores gravitate toward the value of 2.0. This occurs because the block mean difference is divided by half of itself. Indeed, if random noise is eliminated, every person taking the IAT will be characterized as strongly prejudiced, based on the current diagnostic criteria (because an IAT score greater than 0.65 is deemed indicative of a “strong automatic preference for Whites over Blacks”). This will be true whether the difference in the compatible and incompatible response latencies is as small as 1 millisecond or as large as 10,000 milliseconds. In essence, every respondent taking the test begins with a baseline score of strong prejudice (i.e., 2.0 or -2.0); then, their scores move away from this extreme value as a function of their within-block trial error.

Although random noise impacts all attitudinal measures, it typically does so by pushing the true score estimates of some individuals upward and those of others downward. As random error is brought under control, the true scores emerge. By contrast, the presence of random noise in the IAT systematically pushes respondents’ scores in the direction of less prejudice. As error is brought under control, artifactual scores of extreme prejudice emerge. Thus, SD_{WI} and the scoring algorithm associated with it are poor candidates for addressing the known GPS confound. In cognitive psychology, a common way of dealing with this problem is to obtain a direct measure of GPS and then to model the interactive dynamics that are operating. Although such alternative scoring methods have been proposed for the IAT (Blanton et al. 2008b), the general point remains. Researchers

must address the psychometric challenges invoked by their new measures and develop viable psychometric models that justify current research applications (see Borsboom 2006).

In the case of the IAT, the issue of GPS has been skirted by the development of a complex scoring algorithm that ostensibly corrects for GPS bias. Researchers have now moved to using this algorithm to provide people with psychological feedback, a practice that seems to suggest that the field can determine a person's implicit biases from their (transformed) IAT scores. One of the pressing challenges facing the implicit measurement movement is the need to develop means of handling the large differences in the cognitive abilities that can exert strong influences on the implicit responses that ostensibly reflect hidden race attitudes.

Association-strength correlates. Matters of GPS threaten most implicit measures because of the experimental psychologists' heavy reliance on response latencies. Even when the attitude score is not based on response latencies, however, an implicit measure activates some cognitive processes in its attempt to access an underlying attitude. For this reason, inventory-specific features must be considered as well. The IAT measures attitudes by supposedly assessing the strength of associations between racial groups and positive/negative constructs. However, many psychological variables appear to be confounded with association strengths as assessed by the IAT. For instance, De Houwer et al. (2005) found evidence that the similarity between two objects (not just evaluations of them) influences IAT scores. Brendl et al. (2001) found evidence that the familiarity of different IAT stimulus items influences IAT scores. Rothermund & Wentura (2004) have documented a complex "salience asymmetry" that appears to influence association strengths measured by the IAT: Because the "Black" category is less familiar than the "White" category to most white participants, and because "Unpleasant" is less familiar than "Pleasant" to

most people, black and unpleasant stand apart in terms of their salience. This asymmetry, in turn, impacts IAT scores.

Social experiences might also be confounded with IAT scores; these can be conceptualized in terms of both trait and state differences. In terms of momentary state differences, Frantz et al. (2004) demonstrated that anxiety about appearing racist might cause individuals to give responses that are typically interpreted as anti-black. More enduring trait differences were suggested by Arkes & Tetlock (2004), who created a thought experiment that they termed the parable of the two Jesses. Readers were asked to consider two individuals. One, like Jesse Jackson, believes that racial discrimination is an ongoing problem that can be attributed to mistreatment of blacks by society. This individual spends his life trying to effect social change to stamp out racial inequality. The other, like Jesse Helms, believes in individual responsibility and vigorously promotes his view that black Americans must take responsibility for the problems in their communities. Because both individuals spend considerable time reflecting on racial stereotypes and status quo racial differences, both might show strong IAT effects. In support of the model that Arkes & Tetlock propose, researchers have found that sympathy and concern for oppressed groups promotes IAT response patterns that are traditionally interpreted as evidence of more negative implicit evaluations of these same groups (Uhlmann et al. 2006).

Reliability. Although reaction time data have been used for decades in cognitive psychology, the use of such data to index stable individual differences is relatively new and has introduced nontrivial challenges into the field. For current reaction-time indices, a tenth of a second can have a consequential effect on a person's score, and such measurement sensitivity can lead to test unreliability. (We use the term unreliable in the classic test theory sense, to indicate that a test score is prone to random error.) If one assumes that the true underlying construct is

stable over time, then test-retest designs reveal the extent to which a measure is free of random error. Higher correlations imply less susceptibility to random noise; the minimally acceptable psychometric standard is usually 0.70, with 0.90 more desirable for assessment purposes.

For the IAT more generally, Steffens & Buchner (2003) assessed test-retest reliability using a paradigm wherein the IAT retest was an immediate replication of a just-measured implicit attitude by the IAT. The test-retest correlations ranged from 0.50 to 0.62. Greenwald et al. (2006) found that the average test-retest reliability of the IAT across a wide range of studies and topics was 0.56. Cunningham et al. (2001) administered the race IAT across four time periods and obtained an average test-retest reliability over two-week periods of just 0.27. These values are far short of acceptable standards and suggest that the challenge of creating reliable indices may be greater for implicit measures based on reaction times.

A related question is whether implicit attitudes are indeed measuring a construct that is stable. Rather than being a relatively unsound measure of a relatively stable construct, it is possible that the IAT shows low test-retest correlations because it is a relatively sound measure of a relatively unstable construct. Consider mood: Mood changes predictably from situation to situation to such an extent that valid mood measures often show low test-retest reliability. Laboratory studies suggest similar interpretations for the IAT: They indicate that race IAT scores can be systematically influenced by such incidental experiences as exposure to black experimenters (Lowery et al. 2001), rap music (Rudman & Lee 2002), reflection on admired blacks (Dasgupta & Greenwald 2001), and viewing blacks in positive situations (a family barbecue) rather than stereotypic ones (a gang situation) (Wittenbrink et al. 2001). Although researchers typically interpret such effects as evidence that the IAT provides a valid assessment of implicit attitudes, such conclusions only follow if one is willing to view the IAT as a measure that taps a state-like—rather than a trait-

like—construct. Unlike mood, racism has typically been thought of as a stable and enduring characteristic of individuals.

Metric meaning. The IAT is used both to provide individuals with psychological feedback (e.g., “you have a strong automatic preference for Whites over Blacks”) and to estimate the prevalence of psychological states (e.g., “27% of people have a strong bias”). None of the other implicit measures has been used so vigorously, but implicit attitude researchers routinely interpret response patterns in their studies in ways that are consistent with IAT diagnostic conventions. This collective tendency underlies the many claims among social psychologists that most people have a degree of implicit or unconscious biases. This interest in knowing the prevalence of racial bias highlights the need for researchers to attend to more than just the validity and reliability of implicit measures. They must also consider the numbering system used to quantify responses to the measures and whether these metrics are arbitrary or meaningful. Such issues have not been considered in the rush to provide psychological diagnoses and to make claims about the prevalence of implicit bias.

Metric meaning is important, however. When Americans are told about a person who stands 4' 1" tall, most of them have a good sense of this person's height and would consider the individual to be “short.” The numbers used to index height in feet and inches are meaningful to most Americans, as they have extensive experience with them. The numbers generated by psychological tests are rarely as revealing: These numbers only gain meaning as researchers build experience with them and begin to associate specific values with specific outcomes or events. Imagine, for instance, that you are presented with a depression inventory that has a metric ranging from 1 to 25. With no additional information, you would not know the meaning of one score versus another. But if you learned that people who receive a depression score of 15 typically engage in three bouts of uncontrollable crying per day and have four daily

instances of suicidal ideation, then you would gain some sense of the meaning of this value. There is no such contextual information for the IAT. In this sense, the IAT has an arbitrary metric.

A metric is arbitrary when it is not known where a given score on the distribution of scores locates a particular individual on the underlying psychological dimension or how a one-unit change on the observed score reflects the magnitude of change on the underlying dimension (Blanton & Jaccard 2006). An individual presented with a race IAT score of 0.65 does not have the necessary experience with this score to know what to make of it. Nor do any psychologists. However, this value (0.65) was chosen by IAT researchers as the cut-off for placing IAT respondents in the category of “strong automatic preference for Whites over Blacks” (A.G. Greenwald, personal communication). To date, no study has linked specific IAT scores to observable, behavioral outcomes reflective of implicit prejudice. The psychometric challenge for those seeking to make prevalence statements about implicit racial bias is to conduct empirical research that will make the numbers of the IAT metric nonarbitrary.

Bias on the decline? The implications of metric arbitrariness within the IAT are illustrated by a seemingly large shift in the prevalence of implicit prejudice in America. This shift has gone unnoticed by social scientists, largely because it was not reported in scientific journals.

Table 1 presents the frequency distribution for race IAT diagnoses as reported on the IAT website in 2007 and 2002. According to the website, 28% of the thousands of people who took the test in 2007 are characterized as having a “strong automatic preference for Whites over Blacks.” This number stands in stark contrast to the 2002 estimate of 48%. What is the reason for this seeming drop in implicit prejudice? The shift occurred because the architects of the IAT changed the criteria for classifying people into the different preference categories (at the same time they adopted a new scoring algorithm, described above). Whereas

Table 1 Distribution of implicit association test scores by category

Category	2007 (%)	2002 (%)
Strong automatic preference for whites over blacks	27	48
Moderate automatic preference for whites over blacks	27	13
Slight automatic preference for whites over blacks	16	12
No preference	17	12
Slight automatic preference for blacks over whites	6	6
Moderate automatic preference for blacks over whites	4	4
Strong automatic preference for blacks over whites	2	6

the old criteria used a threshold value corresponding to a Cohen’s *d* of 0.80 for placing people into the extreme category, the new criteria uses a Cohen’s *d* value closer to 1.30. The bases for this change in threshold value are unclear. If the original Cohen’s *d* criteria are applied to current data, then instead of a decrease in implicit prejudice, there would be an increase. Specifically, almost 60% of the people taking the IAT would be diagnosed as having a “strong automatic preference for Whites over Blacks.”

Our goal is not to quibble about what percentage of people are or are not prejudiced. Rather, our goal is to call attention to the casualness with which threshold values have been chosen by researchers faced with an arbitrary metric. Given the societal consequences of the labels and their current uses, there is a need for a more cautious interpretation of IAT scores, at least until empirical research can ground this enterprise.

Arbitrary zero points. The race IAT is a measure of the difference between two attitudes, namely attitudes toward white people relative to attitudes toward black people. In this sense, the test combines two different attitudinal constructs. The architects of the IAT assume that a score of zero on the measured IAT maps onto the true zero point of “no preference” on the underlying dimension of relative racial preference. Other implicit measures of attitudes embrace the same scaling assumption, where a scale value of zero represents the dividing line

Arbitrary metric: a measurement system in which the numbers assigned to categories, although possibly reliable and valid, do not allow one to infer a person’s absolute standing on the underlying psychological dimension

between problack and antiblack evaluations (see Dovidio & Gaertner 2004).

Although the zero-point interpretation has intuitive appeal, there is little empirical evidence to support this interpretation. To gain empirical perspective on the IAT zero point, it is necessary to develop a strong theory that makes predictions about how people differ depending on whether they have positive or negative scores. For instance, researchers might test whether an IAT score of zero discriminates among people who show a clear hiring preference for whites over blacks versus those who show a clear hiring preference for blacks over whites. If the IAT zero point does not discriminate between those who show one preference versus the other, then its presumed mapping onto a true neutral point is questionable. To date, the empirical base for zero-point assumptions is lacking. The many confounds described earlier offer reason to expect the zero-point assumption to be untenable. Another research challenge for implicit measures of racism is to validate the presumed zero points of these scales.

Conclusions on the Measurement of Implicit Racial Attitudes

Although our discussion of the IAT may seem harsh, we do not intend to dampen the enthusiasm for implicit measures of attitudes. Rather, it seems to us that in the rush to embrace implicit measures—fueled in part by enthusiasm for their claimed ability to tap the unconscious—that some core psychometric issues have been brushed aside. Implicit measures have been embraced within psychology, and researchers have drawn strong inferences about the nature and prevalence of racism based upon them. We believe that the current enthusiasm for this measurement approach has outpaced its empirical base. There is much more work to be done before implicit measures can claim to live up to the promises made about them. Our review is intended to identify some future directions for this work.

CONCLUDING COMMENTS

Unconscious racism is an intriguing concept that is enjoying widespread attention. Research in this area has leaped forward with the advent of reaction time measures of implicit attitudes. Reaction time data are widely used in cognitive psychology, and this has been the basis of important and elegant theorizing about information processing and the nature of human memory. We have no doubt that many significant behaviors that people perform are based on information from their environments—information that they may not be able to verbalize or articulate. To us, an important area of study for the analysis of racist actions at the individual level is to examine how people process the cues around them to form impressions and give meaning to their environments. Advances in this area will undoubtedly complicate current conceptions of unconscious racism.

However, for the implicit measurement of racism (conscious or unconscious), it seems that caution is warranted. Despite researchers' tendency to interpret implicit race data as evidence that unconscious racism is pervasive, logical analysis of this construct and closer inspection of the properties of implicit measures suggest to us that strong conclusions are not warranted at this time. Aside from the measurement challenges highlighted here, it also is worth noting that virtually all research related to unconscious racism has been conducted in laboratory settings, with either hypothetical scenarios or behaviors that are of marginal relevance to overt discriminatory behavior. It is unclear whether the results obtained in highly controlled laboratory settings, typically using college student samples, can be generalized to the complex, multivariate, real-world dynamics where experience, social norms, and accountability pressures also guide decisions (see Mitchell & Tetlock 2006). Rather than providing a "royal road to the unconscious," some of the research reviewed here may reveal little more than the extent to which unconscious processes can be co-opted by clever experimenters.

SUMMARY POINTS

1. Unconscious racism can refer to one's inability to (*a*) perceive the racial implications of one's action, (*b*) perceive the causes of racially biased behavior, or (*c*) subjectively note one's own racist attitudes.
2. Research on institutional and symbolic racism has provided examples of how actors might fail to see the racial implications of their actions.
3. Research on person perception has revealed how social factors can produce unconscious racially biased behavior.
4. Research strategies for showing that people are not able to subjectively note their own racist attitudes are in development. The most aggressive research line focuses on implicit attitude measures.
5. Implicit attitude measures ostensibly assess unconscious mental processes, but the methods for establishing that a given response reflects the workings of an unconscious attitude are not well developed.
6. The implicit association test (IAT) is the most widely used measure of implicit attitudes, and strong claims have been made about its ability to reveal high rates of unconscious racism. Empirical evidence does not support these claims.

FUTURE ISSUES

1. The challenge facing implicit measures of racism is controlling for conscious processes, environmental influences, and GPS so that these measures' ability to assess unconscious states can be better evaluated.
2. The ability of unconscious constructs to predict discriminatory behavior over and above conscious constructs (where modern-day as opposed to outdated models of explanatory constructs are used) needs to be explored with greater scientific rigor.
3. Experiments are attempting to document the influence of constructs related to unconscious racism in controlled laboratory settings. Future research should explore these issues in real-world settings in which personal experiences, social norms, and accountability pressures might override or interact with unconscious influences.
4. Measures of unconscious racism with nonarbitrary metrics should be developed if researchers wish to ascertain the prevalence of unconscious racism and/or implicit prejudice in society.
5. The mediators and moderators of unconscious influence need to be better defined.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aberson CL, Ettlin TE. 2004. The aversive racism paradigm and responses favoring African Americans: meta-analytic evidence of two types of favoritism. *Soc. Jus. Res.* 17:25–46

- Ajzen I. 2005. *Attitudes, Personality, and Behavior*. Milton Keynes, Engl.: Open Univ. Press/McGraw-Hill. 2nd ed.
- Arkes HR, Tetlock PE. 2004. Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the implicit association test?" *Psychol. Inq.* 15:257–78
- Ayres I. 2001. *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination*. Chicago: Univ. Chicago Press
- Bagenstos SR. 2007. Implicit bias, "science," and antidiscrimination law. *Harv. Law Policy Rev.* 1:477–93
- Banaji MR. 2001. Implicit attitudes can be measured. In *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, ed. HL Roediger III, JS Nairne, I Neath, AM Surprenant, pp. 117–50. Washington, DC: Am. Psychol. Assoc.
- Bargh JA. 2004. Being here now: Is consciousness necessary for human freedom? In *Handbook of Experimental Existential Psychology*, ed. J Greenberg, SL Koole, T Pyszczynski, pp. 385–97. New York: Guilford
- Bargh JA, Pietromonaco P. 1982. Automatic information processing and social perception: the influence of trait information presented outside of conscious awareness on impression formation. *J. Pers. Soc. Psychol.* 43:437–49
- Blanton H, Jaccard J. 2006. Arbitrary metrics in psychology. *Am. Psychol.* 61:27–41
- Blanton H, Jaccard J, Gonzales PM, Christie C. 2006. Decoding the implicit association test: implications for criterion prediction. *J. Exp. Soc. Psychol.* 42:192–212
- Blanton H, Jaccard J, Christie C, Gonzales PM. 2007. Plausible assumptions, questionable assumptions and post hoc rationalizations: Will the real IAT, please stand up? *J. Exp. Soc. Psychol.* 43:399–409
- Blanton H, Jaccard J, Mitchell G, Tetlock PE. 2008a. *Strong claims and weak evidence: reassessing the predictive validity of the race IAT*. Work. Pap., Dep. Psychol., Texas A&M Univ.
- Blanton H, Jaccard J, Mitchell G, Tetlock PE. 2008b. *The problematic D in the IAT*. Work. Pap., Dep. Psychol., Texas A&M Univ.
- Borsboom D. 2006. The attack of the psychometricians. *Psychometrika* 71:425–40
- Brendl CM, Markman AB, Messner C. 2001. How do indirect measures of evaluation work? Evaluating the inference of prejudice in the implicit association test. *J. Pers. Soc. Psychol.* 81:760–73
- Carmichael S, Hamilton C. 1967. *Black Power: The Politics of Liberation in America*. New York: Vintage
- Correll J, Park B, Judd CM, Wittenbrink B, Sadler MS, Keesee T. 2007. Across the thin blue line: police officers and racial bias in the decision to shoot. *J. Pers. Soc. Psychol.* 92:1006–23
- Cunningham WA, Preacher KJ, Banaji MR. 2001. Implicit attitude measures: consistency, stability, and convergent validity. *Psychol. Sci.* 12:163–70
- Dasgupta N, Greenwald AG. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81:800–14
- Dehaene S, Naccache L, Cohen L, Le Bihan D, Mangin JF, et al. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4:752–58
- De Houwer J, Geldof T, De Bruycker E. 2005. The implicit association test as a general measure of similarity. *Can. J. Exp. Psychol.* 59:228–39
- Devine PG. 1989. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56:5–18
- Dovidio JF, Gaertner SL. 2004. Aversive racism. *Adv. Exp. Soc. Psychol.* 37:95–150
- Duncan BL. 1976. Differential social perception and attribution of intergroup violence: testing the lower limits of stereotyping of blacks. *J. Pers. Soc. Psychol.* 34:590–98

- Festinger L. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford Univ. Press
- Fazio RH, Jackson JR, Dunton BC, Williams CJ. 1995. Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *J. Pers. Soc. Psychol.* 69:1013–27
- Fishbein M, Ajzen I. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley
- Frantz CM, Cuddy AJC, Burnett M, Ray H, Hart A. 2004. A threat in the computer: the race implicit association test as a stereotype threat experience. *Pers. Soc. Psychol. Bull.* 30:1611–24
- Freud S. 2002 (1901). *The Psychopathology of Everyday Life*. New York: Penguin-Putnam
- Gaertner SL, Dovidio JF. 1977. The subtlety of white racism, arousal, and helping behavior. *J. Pers. Soc. Psychol.* 35:691–707
- Gallup Org. 2001. *Black-White Relations in the United States: 2001 Update*. Washington, DC: Gallup
- Gawronski B, Bodenhausen GV. 2008. Unraveling the processes underlying evaluation: attitudes from the perspective of the APE model. *Soc. Cogn.* In press
- Gawronski B, Hofmann W, Wilbur CJ. 2006. Are “implicit” attitudes unconscious? *Conscious. Cogn.* 15:485–99
- Gladwell M. 2005. *Blink: The Power of Thinking without Thinking*. New York: Little, Brown
- Greenwald AG, Krieger LH. 2006. Implicit bias: scientific foundations. *Calif. Law Rev.* 94(4):945–67
- Greenwald AG, McGhee DE, Schwartz JLK. 1998. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74:1464–80
- Greenwald AG, Nosek BA, Banaji MR. 2003. Understanding and using the implicit association test. I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85:197–216
- Greenwald AG, Nosek BA, Sriram N. 2006. Consequential validity of the implicit association test: comment on Blanton and Jaccard. *Am. Psychol.* 61:56–61
- Greenwald AG, Poehlman TA, Uhlmann E, Banaji MR. 2008. Understanding and using the implicit association test. III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* In press
- Jaccard J, Blanton H. 2007. A theory of implicit reasoned action: the role of implicit and explicit attitudes in the prediction of behavior. In *Prediction and Change of Health Behavior: Applying the Reasoned Action Approach*, ed. I Ajzen, D Albarracin, J Hornik, pp. 69–81. Mahwah, NJ: Erlbaum
- Kang J. 2005. Trojan horse of race. *Harv. Law Rev.* 118:1489–593
- Kang J, Banaji MR. 2006. Fair measures: a behavioral realist revision of “affirmative action.” *Calif. Law Rev.* 94:1063–118
- Kahneman D, Tversky A. 1973. On the psychology of prediction. *Psychol. Rev.* 80:237–51
- Karpinski A, Hilton JL. 2001. Attitudes and the implicit association test. *J. Pers. Soc. Psychol.* 81:774–88
- Kovel J. 1977. *White Racism: A Psychobiography*. New York: Random House
- Kunda Z, Thagard P. 1996. Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory. *Psychol. Rev.* 103:284–308
- Lepore L, Brown R. 1997. Category and stereotype activation: Is prejudice inevitable? *J. Pers. Soc. Psychol.* 72:275–87
- Lowery BS, Hardin CD, Sinclair S. 2001. Social influence effects on automatic racial prejudice. *J. Pers. Soc. Psychol.* 81:842–55
- McConnell AR, Leibold JM. 2001. Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *J. Exp. Soc. Psychol.* 37:435–42
- Mitchell G, Tetlock PE. 2006. Antidiscrimination law and the perils of mindreading. *Ohio State Law J.* 67:1023–1121

- Nier JA. 2005. How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Proc. Intergroup Rel.* 8:39–52
- Nosek BA, Smyth FL, Hansen JJ, Devos T, Lindner NM, et al. 2007. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18:36–88
- Olson MA, Fazio RH. 2004. Implicit measures in social cognition research: their meaning and use. *Annu. Rev. Psychol.* 54:297–327
- Olson MA, Fazio RH. 2008. Discordant evaluations of blacks affect nonverbal behavior. *Pers. Soc. Psychol. Bull.* 33:1214–24
- Payne BK. 2001. Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *J. Pers. Soc. Psychol.* 81:181–92
- Payne BK, Burkley MA, Stokes MB. 2008. Why do implicit and explicit attitude tests diverge? The role of structural fit. *J. Pers. Soc. Psychol.* 94:16–31
- Perugini M, O’Gorman R, Prestwich A. 2007. An ontological test of the IAT: self-activation can increase predictive validity. *Exp. Psychol.* 54:134–47
- Phelps EA, O’Connor KJ, Cunningham WA, Gatenby JC, Funayama ES, et al. 2000. Amygdala activation predicts performance on indirect measures of racial bias. *J. Cogn. Neurosci.* 12:729–38
- Quillian L. 2006. New approaches to understanding racial prejudice and discrimination. *Annu. Rev. Sociol.* 32:299–328
- Richeson JA, Shelton JN. 2003. When prejudice does not pay: effects of interracial contact on executive function. *Psychol. Sci.* 14:287–90
- Rothermund K, Wentura D. 2004. Underlying processes in the implicit association test: dissociating salience from associations. *J. Exp. Psychol.* 133:139–65
- Rudman LA, Lee MR. 2002. Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Proc. Intergroup Rel.* 5:133–50
- Saujani RM. 2003. “The implicit association test”: a measure of unconscious racism in legislative decision-making. *Mich. J. Race Law* 395:395–424
- Sears DO, Kinder DR. 1971. Racial tensions and voting in Los Angeles. In *Los Angeles: Viability and Prospects for Metropolitan Leadership*, ed. WZ Hirsch. New York: Praeger
- Schacter DL. 1987. Implicit memory: history and current status. *J. Exp. Psychol. Learn. Mem. Cogn.* 13:501–18
- Shelton JN, Richeson JA, Salvatore J, Trawalter S. 2005. Ironic effects of racial bias during interracial interactions. *Psychol. Sci.* 16:397–402
- Shermer M. 2006. He’s a racist. So are you. So am I. *Los Angeles Times*, Nov. 24, Sect. A, p. 41
- Spruyt A, Hermans D, De Houwer J, Eelen P. 2002. On the nature of the affective priming effect: affective priming on naming responses. *Soc. Cogn.* 20(3):227–56
- Steffens MC, Buchner A. 2003. Implicit association test: separating transsituationally stable and variable components of attitudes toward gay men. *Exp. Psychol.* 50:33–48
- Tetlock PE. 1994. Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Polit. Psychol.* 15:509–29
- Uhlmann EL, Brescoll VL, Paluck EL. 2006. Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *J. Exp. Soc. Psychol.* 42:491–99
- Vanman EJ, Saltz JL, Nathan LR, Warren JA. 2004. Racial discrimination by low-prejudiced whites’ facial movements as implicit measures of attitudes related to behavior. *Psychol. Sci.* 15:711–14
- Vedantam S. 2005. See no bias. *Washington Post*, Jan. 23, Sect. W, p. 12

- Wegener DT, Petty RE. 1997. The flexible correction model: the role of naive theories of bias in bias correction. In *Advances in Experimental Social Psychology*, ed. MP Zanna, Vol. 29, pp. 141–208. Mahwah, NJ: Erlbaum
- Wittenbrink B, Judd CM, Park B. 2001. Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *J. Exp. Soc. Psychol.* 37:244–52
- Zajonc RB. 1968. Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol.* 9:1–27
- Ziegert JC, Hanges PJ. 2005. Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *J. Appl. Psychol.* 90:553–62