Psychological Inquiry
2004, Vol. 15, No. 4, 279–310

COMMENTARIES

No Place for Nostalgia in Science: A Response to Arkes and Tetlock

Mahzarin R. Banaji
Department of Psychology
Harvard University

Brian A. Nosek
Department of Psychology
University of Virginia

Anthony G. Greenwald
Department of Psychology
University of Washington

Ask an attitude expert about the major shifts in thinking about the concept of prejudice since 1954, and the answers will likely contain the following assessments about the broad, modern scientific understanding of the concept:

1. Prejudice and other attitudes were assumed to operate largely in conscious (explicit, deliberate, controllable, intentional) mode. Now they are generally viewed as also operating in a less conscious (implicit, spontaneous, uncontrollable, unintentional) mode (Eagly & Chaiken, 1993; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). From this conceptual shift other changes have followed:

   a. Historically, attitudes were almost exclusively assessed through self-report measures. Now, more indirect methods have been added, notably response latencies to object + evaluation pairings (Fazio, et al. 1986). These measures are thought to reveal less accessible, more automatic forms of attitudes.

   b. Explicit and implicit attitudes can be dissociated, such that one form of the attitude can be evaluatively positive, the other negative. For instance, individuals who endorse egalitarian values broadly, and (honestly) endorse favorable social group attitudes, can nonetheless show negativity on implicit measures (Greenwald & Banaji, 1995).

   c. At the same time, explicit and implicit attitudes can be associated such that those individuals who tend to report higher levels of explicit prejudice are also likely to reveal higher levels of implicit prejudice. In the domain of social group attitudes these relations are sometimes observed to be as high as $r = .50$ (Cunningham, Nezlek, & Banaji, in press; Nosek, 2004), and implicit–explicit correlations more generally have been observed to be as high as $r = .86$ (Greenwald, Nosek, & Banaji, 2003). The psychologically and pragmatically interesting cases are those in which a significant correlation still reveals two separate factors at work (Cunningham, Nezlek, et al., in press).

   d. Discriminatory behavior is predicted by both explicit and implicit measures, but prediction by implicit measures tends to be stronger (Poehlman, Uhlmann, Greenwald, & Banaji, 2004).

2. Psychologists used to think of the concept of prejudice as prominently containing the property of *animus* or *antipathy*, but that is no longer assumed to be a necessary condition (Dovidio, Glick, & Rudman, in press; Glick & Fiske, 2001a, 2001b; Jackman, 1994). In a related vein, the constructs of attitude and stereotype were often conflated, as evidenced in the widely shared but incorrect assumption that evaluations of women are negative. Eagly (Eagly & Mladinic, 1989; see also, Eagly & Diekman, in press) corrected this error showing that attitudes toward women are positive even though stereotypes of them in particular roles can be strikingly negative.

3. More generally, human behavior was once regarded as motivated by rational thought, but now many exceptions are recognized (e.g., Kahneman, Slovic, &

Tversky, 1982; Simon, 1983). Computations that underlie social attitudes and judgment, even those that have moral bearing, are no exception (Banaji & Bhaskar, 2000). Thinking in this way demystifies otherwise troublesome concepts like prejudice by placing them squarely within the purview of ordinary cognition.

In this response to Arkes & Tetlock's (this issue) critique, we raise three issues. First, we challenge the notion of attitude and prejudice as constructs that operate only in conscious form. We see no reason for this burden to be borne by some constructs like attitude or prejudice and not by others mental constructs such as attention, perception, and memory. Just as we speak about explicit and implicit memory measures or systems, so might we profitably speak of explicit and implicit attitude measures or systems. In particular, Arkes and Tetlock do not accurately represent the position of those who study implicit social cognition. They invoke an oxymoron by using the term *endorsement* to refer to the workings of implicit, less conscious or controllable, attitudes.

Second, we show that it is not possible to set aside the concept of implicit prejudice by suggesting that it reflects *mere* association—unless Arkes and Tetlock (this issue) wish to admit that mere associations produce convergent (and discriminant) validity with measures of prejudice as well as rapidly emerging data on criterion validity. Finally, in the work of others, the notion of prejudice as antipathy has been broadly challenged, and Arkes and Tetlock questions have the benefit of alerting scholars to the ongoing redefinition of the concept.

Genuine, 100% Prejudice, Please

Greenwald and Banaji (1995) defined implicit attitudes as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feelings toward an attitude object" (p. 6). Arkes and Tetlock (this issue) stated that implicit attitudes are "an attitude one *endorses* at some level" (emphasis added). These two definitions are at odds in one sense, although Greenwald and Banaji would agree that Arkes and Tetlock's definition is a perfectly fine description of the construct of explicit attitude.

The term *endorses* means "to give approval of or support to, especially by public statement" (American Heritage Dictionary, 1992), or "to approve openly; *especially*: to express support or approval of publicly and definitely" (Merriam-Webster's Online Dictionary, 2004; italics in original). Inherently then, *endorsement* is a characteristic of explicitly stated attitudes. Endorsement is not a characteristic of indirect assessment tools—whether it be response latency measures such as the Implicit Association Test (IAT) or evaluative priming (Fazio et al., 1986; Greenwald, McGhee, & Schwartz, 1998), linguistic style (von Hippel, Sekaquaptewa, & Vargas, 1997), assessments of argument quality (Saucier & Miller, 2003), motor (arm flexion) measures (Cacioppo, Priester, & Bernston, 1993), or any of a multitude of other indirect methods. Further, Greenwald and Banaji (1995) reviewed attitude definitions and noted that, even historically, such definitions avoided assumptions of introspective access, awareness, or controllability, suggesting that attitude theorists have always been open to the possibility that attitudes operate at differing levels of consciousness. The historical reliance on self-report measures may have been more from convenience and a lack of alternative measures than a strong theoretical commitment that attitudes operate only as conscious entities. The main point here is that lack of introspective access and lack of conscious control over the contents of consciousness—features that are more characteristic of implicit than explicit attitudes—preclude endorsement. To speak of implicit attitudes as endorsed would be as nonsensical as speaking about a dog endorsing a bone.

A theme that runs through Arkes and Tetlock's (this issue) article has its origins in an article from the mid-1980s (Sniderman & Tetlock, 1986), whose authors criticized the then-emerging notion of a modern, as opposed to old-fashioned, racism because such thinking took away from genuine prejudice—a deep-seated, irrational insistence on the inferiority of Blacks and contempt and hostility and toward them." (p. 186). Almost twenty years later the same idea is expressed when Arkes and Tetlock bemoan the fact that although so much progress has been made in Black emancipation, these social and political changes appear not to be recognized by some social psychologists, including us. Arkes and Tetlock remind readers of the swift and vast progress in American society, that Black and White Americans can now "drink from the same fountain, sleep in the same hotel room, attend the same schools, or intermarry, there is now close to consensus at the level of both mass and elite opinion that de jure segregation is unacceptable." That shift is what is genuine, they say, and that is notable and to be appreciated.

Quaint as these sentiments may sound as markers of progress in twenty-first-century America, the question of social and political progress is neither our expertise nor of relevance to the argument about the nature of attitudes. We only restate our position about the attitude construct, why we believe that differing forms of attitudes are all likely to be genuine, and the evidence that leads us to consider them as unique, but interdependent constructs.

From the earliest days of our work on implicit social cognition, we have taken the position that implicit and explicit attitudes reveal predictive utility in

differing circumstances, a view that naturally flows from the assumption that the two represent psychologically differentiated constructs. A recent meta-analysis by Poehlman et al. (2004) supports this idea by showing that implicit attitudes do not only predict but that they predict better than explicit measures when the target measure is social group discrimination; on the other hand, explicit attitudes predict significantly better than implicit ones when the target objects are consumer items. As such, we have not endorsed the suggestion by Fazio et al. (1995) that automatic attitudes are a "bonafide" pipeline, although we understand the reasoning behind his use of that metaphor. We equally cannot endorse Arkes and Tetlock's (this issue) notion that genuine prejudice is only consciously reportable prejudice, and that it all but vanished when Black Americans were allowed use of all public water fountains.

Attitude measures are keeping pace with advances in technology to allow previously hidden aspects of mental function to be observed, with replication, across laboratories. The resulting phenomena may not always look and feel like their more familiar counterparts, but this cannot be a reason to reject that they exist and have influence. Moving from Newtonian physics to quantum mechanics required large shifts in assumptions, technology, and understanding. There is no reason to assume that the smaller steps in any science that move away from the familiar and comfortable (here, the view of prejudice as only conscious) is any different. To consider only changes in expressed attitude as genuine markers would be no different than arguing that memory as measured by free recall is more genuine than memory revealed by priming. Both are real. Both are genuine.

Although the issue of old versus modern prejudice is addressed by other commentators in this issue, we also speak to it because Arkes and Tetlock's (this issue) point encompasses the work on implicit attitudes with which we are associated in a unique way. Given their position, Arkes and Tetlock's expressed irritation with us is understandable. If the logic underlying the Modern Racism Scale (McConahay, 1986)—to develop scale items that no longer asked about whether drinking fountains should be desegregated but to accommodate to new standards of attitude and behavior—is viewed by Arkes and Tetlock as a step in the wrong direction, getting away from tapping genuine prejudice, then measures of mental speed assessing associations in memory can only signal the apocalypse. This difference is a fundamental one separating us from Arkes and Tetlock. Given their position that (a) genuine attitudes are those that are consciously expressed and (b) that modernized items on self-report measures are not necessarily measures of prejudice, it would be a stretch for Arkes and Tetlock to accept many of the measures of attitude that are now routinely used— priming, linguistic markers, motor responses, and the IAT (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald et al., 1998)— as revealing preferences, attitudes, feelings. It would be akin to asking the Fuller Court of *Plessy v. Ferguson* (1896) to accept the Warren Court's *Brown v. Board of Education* (1954) decision. The Plessey judges (minus Harlan) would indeed be puzzled as to why emancipation after the Civil war was being ignored and why it is that de jure segregation was being viewed as genuine prejudice.

If Arkes and Tetlock's (this issue) point is that explicit and implicit forms of prejudice should not be blurred, we would concur. Explicit prejudice is distinct from implicit prejudice, hence the different terms, with full recognition of the simplification that any such dichotomy imposes (Banaji, 2001a). We also concur with Arkes and Tetlock that "a person can refrain from explicit prejudice despite having implicit prejudice, but this might require a vigilant effort to prevent the implicit prejudice from manifesting itself in overt behavior." One of us (Banaji, 2001b) in fact used *eternal vigilance* as one practical solution to restoring fairness in decision making. Moreover, at the most public venue in which our opinion is expressed, we state in response to FAQ #7 ("If my IAT shows automatic White preference, does that mean that I'm prejudiced?"):

> Answer: This is a very important question. Social psychologists use the word "prejudiced" to describe people who endorse or approve of negative attitudes and discriminatory behavior toward various out-groups. Many people who show automatic White preference on the Black–White IAT are not prejudiced by this definition. These people are apparently able to function in nonprejudiced fashion partly by making active efforts to prevent their automatic White preference from producing discriminatory behavior. However, when they relax these active efforts, these nonprejudiced people may be likely to show discrimination in thought or behavior. The question of relation between implicit and explicit attitudes is of strong interest to social psychologists, several of whom are doing research on that question for race-related attitudes. ("Project Implicit," n.d.)

Despite this clear position, accessible since the Web site's launch in September 1998, Arkes and Tetlock (this issue) repeatedly characterize the authors of IAT and priming research as using their results to brand those who show modal results as guilty of prejudice.

Our conclusion here is to encourage thinking about attitudes as multiply determined and multiply expressed. A long time ago, William James (1902/1958) spoke about layers of consciousness in a manner that suits the present discussion well:

> Our normal waking consciousness, rational consciousness as we call it, is but one special type of conscious-

> ness, whilst all about it, parted from it by the filmiest of screens, there lie potential forms of consciousness entirely different. We may go through life without suspecting their existence; but apply the requisite stimulus, and at a touch they are there in all their completeness, definite types of mentality which probably somewhere have their field of application and adaptation. No account of the universe in its totality can be final which leaves these other forms of consciousness quite discarded. How to regard them is the question,—for they are so discontinuous with ordinary consciousness. Yet they may determine attitudes though they cannot furnish formulas, and open a region though they fail to give a map. At any rate, they forbid a premature closing of our accounts with reality (p. 388).

Evidence accumulated over the last 2 decades shows the manner in which both conscious and unconscious mental states have their fields of application, in attention and perception, in memory and judgment, and in the social manifestations of these processes. The authors of this article are not alone in experiencing personally and understanding professionally the evidence that our own conscious positive attitudes cannot be relied on in all circumstances. That unendorsed and even disapproved of attitudes are ones that exist and can have their field of application is amply demonstrated in psychology broadly speaking (see Fiske, 1998), and we have summarized the validation of the IAT in predicting behavior and correlating with subcortical brain activity known to tap emotion (Cunningham et al., in press; Greenwald & Nosek, 2001; Phelps et al., 2000; Poehlman et al., 2004). Given the evidence, it would be disingenuous, if not in flagrant opposition to the evidence, to hold that if prejudice is not explicitly spoken, it cannot reflect a prejudice.

Some years ago, one of us wrote a chapter to address the questions raised in the early responses to the IAT (Banaji, 2001a). In that article, reasons were offered for calling the empirical phenomena being observed an *implicit attitude*. We argued from first principles that (a) these phenomena fit with definitions of attitude and prejudice, (b) that lessons from research on human memory, indicating a similar progression from thinking about conscious forms to both conscious and unconscious forms of memory could serve as a model, and (c) multiple demonstrations of the construct validity of implicit attitudes are consistent with the notion of attitudes. The arguments offered there still hold, with improvement: There are now many more published instances of construct and criterion validity available in print (see Greenwald & Nosek, 2001; Poehlman et al., 2004).

"Mere" Association

If Arkes and Tetlock (this issue) mean to derogate implicit social cognition research by referring to implicit attitudes as "mere association," then the effect may be quite the opposite because of the many fundamental contributions that are "mere" or "associative"—mere exposure, associative learning in classical conditioning, and so on. If their intention is to indicate that something that is a "mere association" cannot be an attitude, then a reasoned analysis of such a claim must contend with evidence from Dasgupta, McGhee, Greenwald, and Banaji (2000; Dasgupta, Greenwald, & Banaji, in press) showing that "mere familiarity" cannot account for implicit attitudes measured by the IAT. If by mere association Arkes and Tetlock mean that nothing of importance is being measured, we would point to the work of others showing that implicit measures do indeed predict discriminatory behavior. Here, Fazio et al. (1995) led the way by showing that the strength of negativity on the race priming measure predicted nonverbal negativity toward African Americans. Poehlman et al. (2004) present studies that show that the extent of negativity on the IAT predicts a range of behaviors such as unfriendliness toward African Americans and gay men, rating a Black author's essay negatively, selecting a Black partner, willingness to cut the budget for Jewish or Asian student organizations, criminal sentence strength for Hispanics, discriminating against female job applicants, and physical proximity to Black partner. As Poehlman et al. (2004) noted in their review of 86 samples that include validation measures for the IAT, in the context of social group discrimination, implicit attitudes outperform explicit measures in prediction. Data from implicit measures are also consistent with data from explicit measures (Cunningham, Nezlek, et al., in press; Nosek, Banaji, & Greenwald, 2002). Predicted features of attitudes such as attitude strength and self-presentation moderate the relation between implicit and explicit attitudes (Hofmann, DiBartolo, Holaway, Heimberg, 2004; Nosek, 2004).

Ultimately, Arkes and Tetlock's (this issue) view that implicit attitudes are mere associations must address the evidence on construct and criterion validity. Would Arkes and Tetlock disagree that the now classic experiment by Word, Zanna, and Cooper (1974) does not reveal prejudice because the expressions are mere speech, facial and body muscle movements? Certainly no explicit prejudice was expressed by the interviewers who nevertheless discriminated against African Americans. If Arkes and Tetlock accept those data as evidence of prejudice, they would also accept the data we review here. If they do not consider the evidence from Word et al. to be a type of prejudice, then we, along with many other contemporary theorists and their evidence, have been collectively banished.

For the Love of Antipathy

Arkes and Tetlock's (this issue) critique also includes a contested component of the concept of preju-

dice, that is, whether prejudice must involve animus or antipathy. In support of their argument that prejudice (of the genuine variety) must involve animus, Arkes and Tetlock list three definitions that would lead readers to believe that research on this topic died in the mid-1900s. It is true that in midcentury investigators defined prejudice as involving animus, including the influential view of Gordon Allport (1954) that referred to it as "an antipathy based on a faulty and inflexible generalization" (p. 9). But as chapters in a new volume commemorating the fiftieth anniversary of Allport's *The Nature of Prejudice* (Dovidio et. al., in press) tell the story, in that place and time, Allport could not conceive of prejudice without the antipathy component, but that view is not in agreement with the modern stance. Three contributions have been instrumental in bringing about a change in thinking about the animus component and demonstrating yet again that what seems intuitive may not hold up under the lens of new theory and new evidence. All three emerge from observations of gender relations and extensions of those observations to other power relations.

In *The Velvet Glove*, Mary Jackman (1994) provided a sweeping argument for the role of paternalism in gender, class, and race relations in which she argued against the view that intergroup relations in each of these cases is marked by hostility and conflict. Rather, she identifies the "coercive gleam of persuasion" (p. 1) as underlying these major systems of inequality that play out with the consensual involvement of both the dominant and the dominated. Glick and Fiske (2001a, 2001b) showed, via a measure of personality, the component of benevolence as opposed to hostility in thinking about women. With datasets that impressively cover several countries across the world, they showed that "Benevolent Sexism, though a kinder and gentler from of prejudice, is pernicious" (p. 117).

Likewise, Eagly and Mladinic (1989; see Eagly & Diekman, in press) changed the minds of many by pointing out the inaccuracy of the field's assumption that discrimination against women occurs because attitudes toward women were negative (misogyny). In fact, attitudes toward women are overwhelmingly positive. Eagly and colleagues point out that discrimination against women (in spite of positive attitudes) can result particularly when women violate expected social roles. With such analyses, the field has moved beyond the old-fashioned, comfortable, and inaccurate view that prejudice necessitates animus. These three perspectives have had widespread impact and to find them missing from Arkes and Tetlock's (this issue) report is puzzling.

For individual scientists such as Eagly, Fiske, Glick, and Jackman, observations of discrimination in the presence of positive attitudes were pivotal in seeking an understanding of how such psycho-social situations could arise. If manifest hostility and conflict did not seem to accompany broad systems of discrimination, they asked, what might be the psychological states that produce the quiet coercion that maintains the evident inequality? Their core concern is with the presence of discrimination in the absence of antipathy. The logic here bears similarity to arguments offered by some justices like William Bennett, Thurgood Marshall, and Ruth Bader Ginsburg, that solutions to discrimination should be guided by assessing their impact rather than the explicit intent to harm. If a policy is demonstrably discriminating in its impact (positive or negative) on social groups—that is to say, it produces disparate impact—that ought to serve as the basis of remedies. Other justices such as William Rehnquist and Antonin Scalia have argued that disparate impact is not the way, but that explicit, or genuine prejudice as they may even say, must be demonstrated to redress harm. We take no direct position on this issue here, although it is clear that the work reviewed previously would caution against assuming that harm can only be computed based on the presence of antipathy. Our sense from Arkes and Tetlock's (this issue) positions is that there would be no place for paternalism in their concept of prejudice and that they would side with those justices who demand evidence of genuine prejudice. This is a difference of opinion among us and Arkes and Tetlock, much as it is a difference of opinion among other groups of colleagues such as the justices of the U.S. Supreme Court.

It Wasn't Me

Arkes and Tetlock (this issue) are not alone in struggling with the question of how and where to locate implicit attitudes. Others such as Karpinksi and Hilton (2001) and Olson and Fazio (2004) reported similar worries indicating at the very least that understanding the locus of implicit attitudes is a difficult issue. The problem can be stated thus: Implicit attitudes—and in Olson and Fazio's case, implicit attitudes as measured by the IAT specifically—are not measures of attitude per se; that is, they are not measures of the person's own attitude but rather the person's knowledge of the environment (i.e., something about the culture locally or globally). We understand the urge to create distance from data that do not paint a pretty picture of ourselves, and because we have spoken about this issue before, we restate the position expressed (even endorsed!) by Banaji (2001a):

> The finding of a pro-White effect among White Americans has persistently raised the possibility that what the IAT detects is not a reflection of the individual's own implicit attitude, but rather a preference that resides in some clearly separable culture out there. Culture is offered both as the origin of the automatic preference, the font of the pro-White bias. But further,

some cultural attitude is also what the IAT is assumed to be measuring. I regard the first part of this assessment to be true and the second to be false in a particular sense. It is true that the IAT reflects a learned preference in the same way in which other types of learning reflect the influence of culture—there is, in that sense, nothing special about it. For example, a semantic priming task roughly detects repeated cultural pairing (moderated through individual experience) of say the concepts *doctor* and *nurse.* In the same way, the IAT roughly detects repeated cultural pairing (moderated through individual experience), of Black + bad/White + good, most clearly among non-Black inhabitants of the United States. But just as the strength of association between *doctor* and *nurse* in a given person reflects how those constructs have come to be paired in the mind of a particular individual, so does variation in pro-White bias reflect the strength of association between White + good in an individual mind, however culturally "caused."

The following example should clarify the reason for the mistaken belief that the preference being measured has little to do with an individual's preference. It should surprise no one when we say that it is through cultural learning that children in South India learn to eat and love very hot pickles (even though all infants, including South Indian ones, spit them out with vigor). What is interesting is "whose" attitude toward pickles we then believe the eventual adult attitude to be. I'd argue that we see this attitude as belonging to the individual (i.e., as Suparna's attitude, or Kavitha's attitude), however obvious may be the cultural influence. As a field, we believe, that attitudes, although showing cultural variation (e.g., some Americans liken the taste of Indian pickles to that of gasoline, whereas millions of Indians can't get through a meal without them), also reflects the attitude of the individual embedded in that culture. And to social psychologists, it is the individual differences in those attitudes that are important and interesting, in addition to group differences. Indeed, it is individual variability that is at the core of the construct of attitude.

But why is there such a compelling sense that the implicit attitude that is being picked up is not one's own? The fallacy may arise from assuming a bright line separating self from culture, an assumption that is becoming less tenable as we discover the deep reach of culture into individual minds (Fiske, Kitayama, Markus, & Nisbett, 1998). Implicit attitudes, as I see it, reflect traces of experiences *within* a culture that have become so integral a part of the individual's own mental and social make-up that it is artificial, if not patently odd, to separate such attitudes into "culture" versus "self" parts.

But the more important observation here may be this: The experience that implicit attitudes, as measured by the IAT, may not reflect an individual's own attitude but rather that of the culture may lie in the dominant popular understanding of attitudes—as things that are under conscious awareness, intention, and control. And this is a meaningful experience and distinction that consciously held attitudes certainly allow. That is, one can consciously have the compelling experience of holding a belief or attitude that is discrepant with those of individual others (e.g., "My senator likes the NRA, but I don't") or beliefs that are discrepant from a culture, or subculture (e.g., "97% of all Americans (and 95% of physicians) believe in God, but I don't"). The human ability to consciously "know" one's own attitude or belief, and to "know" its separation from the attitudes and beliefs of others, is an important marker of conscious social cognition. The ability to be able to consciously reflect on one's own mind, a fundamentally unique human ability, is what appears to be causing the confusion regarding implicit attitudes. We desire to see a separation between culture and person in the same way with implicit attitudes as we do with explicit attitudes and we impose this distinction on the data, so powerful is the assumption of individual-culture separation (for a clear example of this fallacy, see Karpinski & Hilton, 1999).[1] The expectation is that just as conscious attitudes are malleable by volition, so must be the case with automatic attitudes. When implicit attitudes do not respond to the call of free will, the source of the attitude becomes suspect—*whose* attitude is it? "Not mine," is the answer, "I can't seem to control it, and surely if it were *mine*, I would be able to do so." Add to this the unpalatable nature of the observed dissociation between conscious and unconscious race attitudes, and we may see why a manufactured distinction between self and culture can seem so compelling, even if incorrect.

Perhaps the struggle to find a place to point the finger, to take the burden of possession off one's self, comes from the inherently political nature of such assessments. We certainly don't see the same agitation when we can't seem to remember a list of words for which we show intact priming. Individuals are the transducers of cultural experience—they provide the physical, social, and psychological shell through which culture speaks. Yet when revealed attitudes are not palatable, the reaction is to look for an answer elsewhere, and pointing to culture (not as the environment in which the attitude is learned, but rather as the "thing" whose attitude is being measured), is perfectly understandable and perfectly wrong." (pp. 138–141)

Recently, this debate has moved to the empirical arena where versions of tests are used that are allegedly more or less likely to tap personal attitudes or cultural associations (Nosek & Hansen, 2004; Olson & Fazio, in press) but our basic point remains that it is less sensible

[1]The manuscript by Karpinski and Hilton (1999) cited here refers to a circulated manuscript prior to publication of their paper. The citation may not exactly match their views as they finally appeared in print.

to think of an sharp line between person and culture when thinking about implicit cognition. We reiterate the point that if such associations did not reflect an attitude it would fail to produce the correlations it does with behavior. Most recently, we have observed strong correlations between IAT measures of race bias and degree of spontaneous smiling to black versus white targets (Olson, Carney, & Banaji, 2004). Such relationships would be hard to explain based on the claim that what such measures detect is knowledge of the culture rather than one's attitude.

Nineteenth-century Rationality?

Given the many meanings of the term *rational* and the complexity of the issues, it is not practical to give this issue the attention that it deserves here, other than offer a few observations. It is possible that Arkes and Tetlock (this issue) make the mistake of conflating *reasonable* with *rational*. If so, there may be no debate here. We borrow directly the arguments offered by Banaji and Bhaskar (1999) about the meanings of rationality as used in contexts such as Arkes and Tetlock's critique and its application to understanding the role of using group knowledge in assessments of individuals.[2]

When stereotypes are unconsciously activated and relied on, there are two direct challenges to the implementation of fairness that are posed: (a) Perceivers and targets are unaware of the rendering of consequential judgments that affect the lives of both, and (b) the decision involves knowledge about the social group rather than the targets alone. These two concerns raise issues of fairness are not inventions of modern, 20th-century concepts of justice. It is a fundamental principle of justice, now almost a thousand years old in Anglo-American jurisprudence (Assize of Clarendon, 1166; Plucknett, 1956), that individuals should be cognizant of the charges against them so as to ensure that judgments are not based on factual error, although a deeper principle is also involved, that justice is better served when an opportunity to be heard exists (Ptahotep scrolls, 2400 B.C.). Judges who are unaware subvert this principle because those who are judged under these circumstances are denied the opportunity to contest, contradict, or modify the judgment.

It is an equally hoary and fundamental principle of justice that judgments about individuals must be based on individuals' own behavior, not those of others who are related to them in any way. Societies in which punishment was based on association (e.g., when families of traitors were beheaded in 17th-century T'ang China) are regarded by the standards of contemporary democracies to be barbaric. In this century, social science research in which beliefs about groups have been shown to influence judgments of individuals has been increasingly interpreted as representing bias. This interpretation arises not from a concern with the correctness of perceivers' beliefs about the group, but because the application of group level knowledge (Some X are Y) to individuals (X is Y) is deemed to be wrong.

If the task is to identify criminals, a guilt-by-association position holds that the greater identification of Black than White is rational and defensible on the basis of base-rate information. On the other hand, many personal and social codes of ethics hold that judgments about individuals should be based on an individual's own behavior without attention to group membership (guilt-by-behavior position). According to this position, it is implausible or incorrect to infer that the parents of murderers are more likely to be murderers because they belong to the same social group (i.e., family) or that because police officers are convicted of crimes at a higher rate than the population (Uviller, 1996), that Officer X is a criminal. This belief that guilt-by-association is morally repugnant is so fundamental that it occupies a central place in all codes of justice from Ptahotep (Ptahotep, 2300 B.C.) to Hammurabi to Asoka (259 B.C., see Nikam & McKeon, 1958) to the Assize of Clarendon (1166; see Plucknett, 1956) to all modern constitutions (with a small number of European exceptions in this century). These general principles provide relevant context for considering the so-called rationality of stereotypes.

Not Classically Rational

Let us say that the task of the subject is to identify names of criminals given a list of names that imply ethnicity. Arkes and Tetlock's (this issue) view is that a reliance on race to make such a decision is simply rational. Following nearly fifty years of research in psychology, we show that the behavior of participants performing such a task does not adhere to classical rationality. Table 1 illustrates a partial list of the many possible utility functions that participants might choose (if they were rational), and an inspection of these suggests why any of them are unlikely descriptors of behavior. Not only do the utility functions require computations that are too complex for subjects unequipped with a calculator to perform, they also require data that even subjects keenly aware of the domain are unlikely to have (e.g., relative frequency of Blacks and Whites in America as a whole, of Blacks and Whites convicted of crimes, of arrested Blacks and Whites, of incarcerated Blacks and Whites, of Black and White names in news reports, number of Type I and Type II errors in news reports, etc.). We do

[2]This section is directly borrowed from a previous article (Banaji & Bhaskar, 2000). Because in some cases the language is edited or slightly changed, we cannot attribute exact quotation. However, we note that this material is not original to this article.

Table 1. *Possible utility functions for participants in race/criminality experiments*

Minimize [(Black names/White names)$_{sample}$ - (Black names/White names)$_{population}$]
Minimize [(Black names/White names)$_{sample}$ - (Black names/White names)$_{arrested}$]
Minimize [(Black names/White names)$_{sample}$ - (Black names/White names)$_{convicted}$]
Minimize [(Black names/White names)$_{sample}$ - (Black names/White names)$_{incarcerated}$]
Minimize [(criminal proportion)$_{sample}$ - (criminal proportion)$_{population}$]
Minimize [(criminal proportion)$_{sample}$ - (criminal proportion)$_{arrested}$]
Minimize [(criminal proportion)$_{sample}$ - (criminal proportion)$_{convicted}$]
Minimize [(criminal proportion)$_{sample}$ - (criminal proportion)$_{incarcerated}$]

Notes. Utility functions 1 through 4 are race-conscious utility functions. Utility functions 5 through 8 are race-neutral. All the utility functions require awareness of the properties of names in the general population, such as the absolute and relative numbers of criminals and non-criminals, and so on. Each of the utility functions also requires a participant to decide how many names to circle based on these ratios, using other criteria that are extrinsic to the problem representation such as which of the particular names to select given the numerical outcome of a utility function.

not dwell on this argument, its conclusions fortunately being in tune with decades of research showing that human behavior is not classically or axiomatically rational (Tversky & Kahneman, 1974; March & Simon, 1958; Newell & Simon, 1972; Simon, 1947, 1955, 1983).

Other Standards for Judgment

Disciplines vary in their methods for determining error. We broadly define four criteria to show that the behavior of using knowledge about the group (however correct it may be) to make judgments about individual members is best characterized as erroneous: *universality of social practice, logic, intention, and analogy*. Because of its most direct relevance, the first is given the most attention. The other three are briefly mentioned and are discussed in greater detail in Banaji and Bhaskar (2000).

Social practice across time and culture has universally recognized the moral discomfort inherent in category-based social judgments. In the last century, Justice Harlan's dissent in *Plessy v. Ferguson* (1896), among the most cited opinions of the Supreme Court, states eloquently that category-based judgments involving race are immoral and cannot be the basis of public policy. In his dissent, he wrote:

> Our constitution is color-blind, and neither knows nor tolerates classes among citizens. … The law regards man as man, and takes no account of his surroundings or of his color when his civil rights as guaranteed by the supreme law of the land are involved. It is therefore to be regretted that this high tribunal, the final expositor of the fundamental law of the land, has reached the conclusion that it is competent for a state to regulate the enjoyment by citizens of their civil rights solely upon the basis of race. In my opinion, the judgment this day rendered will, in time, prove to be quite as pernicious as the decision made by this tribunal in the Dred Scott Case.

American history since has revealed the majority opinion's moral bankruptcy, but we cite Justice Harlan here to ask whether what appeared distasteful in 1897 for public policy might seem unacceptable now for interpersonal and intergroup social judgments.

In the first half of this century, Walter Lippmann (1922/1934) and Gordon Allport (1954) both emphasized the ordinary cognitive bases of category-based judgments, and yet their writings clearly reveal their recognition of the failures inherent in such judgments. Most poignantly, Gunnar Myrdal (1944) showed that Americans experience a moral dilemma "*an ever-raging conflict between, on the one hand, the valuations preserved on the general plane which we shall call the 'American Creed,' where the American thinks, talks, and acts under the influence of high national and Christian precepts, and on the other hand, … group prejudice against particular persons or types of people … dominate his outlook*" (p. xlvii). A half century later, Devine's (Devine, Monteith, Zuwerink, & Elliot, 1991; Zuwerink, Devine, Monteith, & Cook, 1996) work strikingly shows the continued existence of the moral dilemma in the form of heightened guilt among American students confronting their prejudice.

When stating a stereotype in the form of a logical proposition, the appropriate logical quantifier is *some*, *several*, *many*, *a few*, but almost never *all*. The type of logical deduction revealed by experimental participants is of the following kind: "Some members of the set × have characteristic Ω. Object #<22310> is a member of the set ×. Therefore object #<22310> has characteristic Ω." To confuse the logical quantifier *some* with the logical quantifier *all* in the first statement is the kind of error known in logic as a confinement law error (Kalish & Montague, 1964), or in psychology the "atmosphere effect" (Woodworth & Sells, 1935). Premises containing *some* create an atmosphere for accepting inferences that actually deserve the answer "can't say—*no specific* conclusion follows from the premises. If a person accepts a specific conclusion for an invalid syllogism, that is an error in reasoning, and such errors frequently conform to

predictions based on the atmosphere hypothesis" (Bourne, Dominowski, & Loftus, 1979, p. 277).

In a different approach, for many circumstances an outcome is considered incorrect if it is inconsistent with one's *intention*. Intending to drive on the right side of a road, but ending up on the left is an error. In a similar way, intending to feel and behave in line with one's values, but failing to do so can be considered an error. In fact, recognizing the inconsistencies between *ought* and *actual* is apparently what accounts for the discomfort expressed when a mismatch between desired feelings and behaviors versus actual feelings and behavior are highlighted (Devine et al. 1991). How a society should choose to deal with such errors and their consequences is a separate question and one that is beyond the scope of this article. Our purpose is to emphasize that conclusions about decision making that are disturbing ought not to be mischaracterized as benign or correct.

A final argument for considering experimental results as representing error can be made by analogy. In other areas in which similar criteria of incorrectness as in our experiments are met, the behavior is routinely classified as an error. For example, when two objects that are identical in shape and size (such as table tops in Shepard's, 1990, p. 48, parallelogram illusion) are perceived to be dissimilar, we regard the resulting misperception to be a remarkable error. Explanations concerning the origin of the perceptual error do not produce a desire to recategorize the error as reflecting a correct judgment. Likewise, when two behaviors are identical (one performed by Malik, the other by Mark) but are not judged to be so, we must regard the resulting misperception to be an error. The confusion created about whether to regard the latter example as an error compared to the former that obviously is, may most charitably be understood as reflecting a desire to avoid confronting the seamy side of decision making that accompanies such social judgments.

Conclusion

Throughout the critique, Arkes and Tetlock's (this issue) arguments rely on earlier modes of thinking about attitude and prejudice. This is evinced in their difficulty with the modern notion that conscious prejudice is but one form of prejudice, in ignoring evidence about implicit attitude validity by referring to the concept as reflecting "mere association," in setting aside the work of social scientists more broadly who have argued that prejudice need not involve antipathy and by confusing reasonableness with rationality. In so doing, their views do not match modern conceptions of attitude and prejudice. More problematic, Arkes and Tetlock's arguments are inconsistent with the large contemporary body of evidence on attitudes and prejudice.

Notes

This work was supported by grants from the National Science Foundation and the National Institute of Mental Health to M. Banaji, A. G. Greenwald, and B. A. Nosek. It was also supported by a grant from the Third Millennium Foundation and a fellowship from the Rockefeller Foundation to M. R. Banaji.

We thank Roy Ruhling and Winmar Way for help in preparing the manuscript and Dana Carney, Dolly Chugh, Jeff Ebert, and Jason Mitchell for their thoughtful comments on the manuscript.

Mahzarin R. Banaji, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: mahzarin_banaji@harvard.edu

References

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

American Heritage Dictionary. (1992). 3rd ed. New York: Houghton Mifflin.

Banaji, M. R. (2001a). Implicit attitudes can be measured. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banaji, M. R. (2001b, January). *Paradoxes of mind and society.* Paper presented at DeVane Lecture, Democratic Vistas, Tercentennial Lectures, Yale University, New Haven, CT.

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139–175) Cambridge, MA: Harvard University Press.

Bourne, L. E., Dominowski, R. L., & Loftus, E. (1979). *Cognitive processes*. Englewood Cliffs, NJ: Prentice Hall.

Brown v. Board of Education, 347 U.S. 483 (1954).

Cacioppo, J. T., Priester, J. R., & Bernston, G. G. (1993). Rudimentary determination of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65,* 5–17.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (in press). Dissociated conscious and unconscious evaluations of social groups: An fMRI investigation. *Psychological Science*.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (in press). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*.

Dasgupta, N., Greenwald, A. G., Banaji, M. R. (in press). The first ontological challenge to the IAT: Attitude or mere familiarity? *Psychological Inquiry.*

Dasgupta, N., McGhee, D. E, Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental and Social Psychology, 36,* 316–228.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality & Social Psychology, 60,* 817–830.

Dovidio, J. F, Glick, P., & Rudman, L. (Eds.). (in press). *On the nature of prejudice: 50 years after Allport.* Malden, MA: Blackwell.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich.

Eagly, A. H., & Diekman, A. B. (in press). What is the problem? Prejudice as an attitude-in-context. In J. F. Dovidio, P. Glick, & L. Rudman (Eds.), *Reflecting on the nature of prejudice*. London, UK: Blackwell.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin, 15,* 543–558.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality & Social Psychology, 69,* 1013–1027.

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

Fiske, A., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 915–981). San Francisco: McGraw-Hill.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). New York: McGraw-Hill.

Glick, P., & Fiske, S. T. (2001a). Ambivalent alliances: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *55,* 109–118.

Glick, P., & Fiske, S. T. (2001b). Ambivalent sexism. In M. P. Zanna (Ed.), *Advances in experimental social psychology, 33,* 115–188.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review. 102,* 4–27.

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. & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie*, *48,* 85–93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: 1. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2004). *A meta-analysis on the correlation between the Implicit Association Test and explict self-report measures.* Unpublished manuscript. University of Trier, Germany.

Jackman, M. R. (1994*). The velvet glove: Paternalism and conflict in gender, class, and race relations.* Berkeley: University of California Press.

James, W. (1958). *The varieties of religious experience*. New York: New American Library. (Original work published 1902)

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kalish, D., & Montague, R. (1964). *Logic; techniques of formal reasoning.* New York: Harcourt Brace.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788.

Lippmann, W. (1934). *Public Opinion*. New York: Macmillan. (Original work published 1922)

March, J. G., & Simon, H. A. (1958). *Organizations.* New York: Wiley.

McConahay, J. P. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic.

Merriam-Webster's online dictionary. (2004). Retrieved September 21, 2004 from http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=endorses

Myrdal, G. (1944). *An American dilemma: The Negro problem in modern democracy.* New York: Harper.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Nikam, N. A., & McKeon, R. (1958). *Asoka, King of Magadha.* Chicago: University of Chicago Press.

Nikam, N. A., & McKeon, R. (Eds.) (1959). *The Edicts of Asoka.* Chicago: University of Chicago Press.

Nosek, B. A. (2004). *Moderators of the relationship between implicit and explicit attitudes.* Unpublished manuscript.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dynamics, 6, 1,* 101–115.

Nosek, B. A., & Hansen, J. (2004). *The associations in our heads belong to us: Measuring the multifacted attitude construct in implicit social cognition.* Unpublished manuscript, University of Virginia.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86,* 653–667.

Olson, K. R., Carney, D., & Banaji, M. R. (2004). Implicit attitudes predict facial mimicry. Abstract submitted for publication.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, S., Gatenby, J. C. Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729–738.

Plessy v. Ferguson. *163,* U.S. 537 (1896).

Plucknett, T. F. T. (1956). *A concise history of the common law*. Boston: Little, Brown & Co.

Poehlman, A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2004). *Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.* Unpublished manuscript, Yale University, New Haven, CT.

Project Implicit. (n.d.) Retrieved September 23, 2004, from https://implicit.harvard.edu/implicit/

Ptahotep. (2300 B.C.). *The Instruction of Ptahotep* (6th Dynasty). 2300–2150 B.C.

Saucier, D. A**.**, & Miller, C. T. (2003). The persuasiveness of racial arguments as a subtle measure of racism. *Personality and Social Psychology Bulletin, 29,* 1303–1315.

Shepard, R. N. (1990). *Mind sights: Original visual illusions, ambiguities, and other anomalies.* New York; Harper & Row.

Simon, H. A. (1947). *Administrative behavior*. New York: Macmillan.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118.

Simon, H. A. (1983). *Reason in human affairs*. Stanford, CA: Stanford University Press.

Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues, 42,* 129–150.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainly: Heurisitcs and biases. *Science, 185,* 1124–1131.

Uviller, R. H. (1996). *Virtual justice: The flawed prosecution of crime in America*. New Haven, CT: Yale University Press.

von Hippel, W., Sekaquaptewa, D. J., & Vargas, P. (1997).The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33,* 490–509.

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

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18,* 451–460

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10,* 109–120.

Zuwerink, M. J., Devine, P. G., Monteith, M. J., & Cook, D. A. (1996). Prejudice toward Blacks: With and without compunction? *Basic and Applied Social Psychology, 18,* 131–150.

Bias on Prejudice? The Politics of Research on Racial Prejudice

Richard E. Redding
Villanova University School of Law
and
Department of Psychology
Drexel University

Ever since the *Scientific American* studies reported dramatic declines in racial prejudice during the 1960s and 1970s (see Sniderman, Piazza, & Harvey, 1998), social psychologists and political scientists have made it their project to discover if racism is endemic in places outside the American consciousness. Claiming that overt racism had "drastically diminished and in some cases has virtually disappeared" (Kinder, 1986, p. 152), they espoused theories of a "new racism," arguing that the express racism of past years had been transformed into an unspoken, even unconscious prejudice. Rather than saying they disliked African Americans, many Whites were "symbolic racists" by supporting conservative views on public policy issues like welfare, affirmative action, and crime that reflected their underlying prejudices about race (Sniderman et al., 1998). Enter research on "implicit prejudice," the next logical step in the progression of the new racism agenda, with leading implicit prejudice researchers espousing symbolic racism theory (see Dovidio, Mann, & Gaertner, 1989).

Psychologists developed new techniques to uncover the new racism. The affective priming technique and the Implicit Association Test (IAT), designed to uncover implicit biases, are well described by Arkes and Tetlock (this issue). Both techniques assume that the faster the reaction time between stimuli, the greater the association between them in semantic memory. The affective priming method involves the presentation of a series of priming and target stimuli and measuring the reaction time for making a judgment about each target stimuli as a function of the prime. Using African-American and White faces as the priming stimuli, Fazio, Jackson, Dunton, and Williams (1995) presented positive and negative adjectives (e.g., *wonderful*, *annoying*) as target stimuli, for which participants were required to respond as quickly as possible by pressing a key labeled *good* or *bad*. White participants had faster reaction times to the good words after being primed with the White faces and faster reaction times to the bad words after being primed with African-American faces. The IAT uses a dual categorization task rather than priming stimuli, comparing reaction times in "compatible" versus "incompatible" conditions. In the compatible condition, participants must respond by pressing one key when a White name/positive word pair is presented and another key when an African-American name/negative word pair is presented. In the incompatible condition, participants must respond when a White name/negative word or an African-American name/positive word is paired. Studies typically find that Whites react faster in the compatible condition and slower in the incompatible condition.

Results such as these have been used to demonstrate the widespread existence of racial prejudice in present-day America and to argue the need for particular social policies, including affirmative action. People are urged to take a Web-based IAT to foster an awareness of their own unconscious racial biases (see Monteith, Voils, & Ashburn-Nardo, 2001). "The results of over one million tests show that unconscious bias exists in most of us" (Tolerance.org, n.d.), reflecting deeply held racist attitudes that are more insidious than explicit attitudes because they may be unconscious yet automatically activated, particularly in situations where race is salient:

> The argument that racial prejudice has "gone underground" has gained a new layer of meaning during the last decade … prejudiced feelings and beliefs continue to be prevalent, but people are willing to express them only indirectly and in relatively subtle ways … with little intent or conscious awareness, negative racial associations that are consciously disavowed can be activated and used as a basis for responding to members of stereotyped groups. This is a particularly dangerous form of prejudice, as even people with the best of intentions may experience great difficulty when trying to avoid [prejudicial] responses. (Monteith et al., 2001, pp. 395–396).