

This article was downloaded by: [University of Florida]

On: 15 January 2010

Access details: Access Details: [subscription number 907139056]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

**PSYCHOLOGICAL  
INQUIRY**

*An International Journal for the  
Advancement of Psychological Theory*

Volume 15, Number 1 Leonard L. Martin & Ralph S. Swann, Editors January-March 2009

 Psychology Press  
Taylor & Francis Group

## Psychological Inquiry

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775648164>

### COMMENTARIES: Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"

Online publication date: 19 November 2009

**To cite this Article** (2004) 'COMMENTARIES: Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"', *Psychological Inquiry*, 15: 4, 279 — 310

**To link to this Article:** DOI: 10.1207/s15327965pli1504\_02

**URL:** [http://dx.doi.org/10.1207/s15327965pli1504\\_02](http://dx.doi.org/10.1207/s15327965pli1504_02)

## PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

---

## 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 & Diekmann, 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 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 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 im-

PLICIT 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 non-verbal 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 form of prejudice, is pernicious” (p. 117).

Likewise, Eagly and Mladinic (1989; see Eagly & Diekmann, 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).<sup>1</sup> 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

<sup>1</sup>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.<sup>2</sup>

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

<sup>2</sup>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

---

<i>Minimize</i> [(Black names/White names) <sub>sample</sub> - (Black names/White names) <sub>population</sub> ]
<i>Minimize</i> [(Black names/White names) <sub>sample</sub> - (Black names/White names) <sub>arrested</sub> ]
<i>Minimize</i> [(Black names/White names) <sub>sample</sub> - (Black names/White names) <sub>convicted</sub> ]
<i>Minimize</i> [(Black names/White names) <sub>sample</sub> - (Black names/White names) <sub>incarcerated</sub> ]
<i>Minimize</i> [(criminal proportion) <sub>sample</sub> - (criminal proportion) <sub>population</sub> ]
<i>Minimize</i> [(criminal proportion) <sub>sample</sub> - (criminal proportion) <sub>arrested</sub> ]
<i>Minimize</i> [(criminal proportion) <sub>sample</sub> - (criminal proportion) <sub>convicted</sub> ]
<i>Minimize</i> [(criminal proportion) <sub>sample</sub> - (criminal proportion) <sub>incarcerated</sub> ]

---

*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  $\times$  have characteristic  $\Omega$ . Object #<22310> is a member of the set  $\times$ . Therefore object #<22310> has characteristic  $\Omega$ ." 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, Tercennial 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., & Diekmann, 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 unobtrusive 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: I. 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 explicit 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 multifaceted 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 uncertainty: Heuristics 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).



Such implicit attitudes, reflecting the “mental residues of a racist culture” (Greenwald & Nosek, 2001, p. 86), are said to drive racist and discriminatory behavior towards minorities.

Without question, biases influence judgment and behavior outside of our conscious awareness. But do the findings of implicit prejudice studies reflect racial prejudice or only cultural stereotypes? Although not entirely new (see Karpinski & Hilton, 2001; Zuriff, 2002), Arkes and Tetlock (this issue) provide a tour-de-force critique of the implicit prejudice literature by challenging the assumption that the IAT and affective priming techniques tap personally held prejudicial attitudes rather than knowledge of well-learned cultural stereotypes. The distinction between stereotype and prejudice is a critical one. A prejudice goes well beyond a stereotype to include hostility and ill-will toward the stereotyped group, irrational prejudgment, an unwillingness to modify one’s stereotypical beliefs based on new or contrary information, and discriminatory beliefs or practices (Sniderman et al., 1998). One may be quite knowledgeable of racial stereotypes, yet be unprejudiced. Indeed, it would be difficult for anyone exposed to American society not to have learned well the cultural stereotypes relating to race. National survey data show that awareness of negative stereotypes of African Americans remains quite common (Carmines & Layman, 1998; Peffley & Hurwitz, 1998).

There is substantial empirical data to support the claim that measures of implicit prejudice may only be measures of stereotype knowledge. First, the average correlation across studies between implicit prejudice and attitudinal measures is relatively low ( $r = .24$ ); moreover, both high and low scorers on explicit attitudinal measures score similarly on measures of implicit prejudice. Taken together, these findings suggest that knowledge of cultural stereotypes is not coextensive with prejudicial attitudes and behavior (Arkes & Tetlock, this issue). Karpinski and Hilton’s (2001) findings are especially compelling: Participant’s IAT scores shifted when presented with new associations between the attitudinal stimuli, but their explicit attitudes remained unchanged. Thus, “IAT scores may reveal little about a person’s beliefs and much about his or her environment or culture ... a culture in which Blacks are devalued relative to Whites. Given that we live in such a culture, it is not surprising that most Americans, White or Black, show a White IAT bias” (Karpinski & Hilton, 2001, p. 786). Second, African Americans show similar IAT responses, suggesting that they also are aware of racial stereotypes, although it is unlikely that they are prejudiced against their own race (Arkes & Tetlock, this issue). (Problematic for this argument, however, are findings that lower status minorities show greater preferences for the majority group than do higher status minorities. Thus, Afri-

can-Americans IAT responses may reflect a rationalization of their lower socioeconomic status by internalizing negative stereotypes of their own racial group; see Rudman, 2004). Finally, many studies rely on relative differences in reaction times, but Arkes and Tetlock point out that having comparatively better attitudes toward one’s own race is not the same as being prejudiced against other races. All racial groups apparently have an ingroup bias (see Perdue, Dovidio, Gurtman, & Tyler, 1990), and Judd, Park, Ryan, Brauer, and Kraus (1995) found greater ethnocentric bias among African Americans than among Whites.

But as further evidence that implicit prejudice measures really do tap prejudice, researchers point out that IAT scores predict Whites’ affective, interpersonal, and nonverbal responses to African Americans (see Arkes & Tetlock, this issue), often better than explicit attitudinal measures (see Montiel et al., 2001; Rudman, 2004). Yet perceived attitude dissimilarity (rather than racial prejudice per se) may mediate the relation between stereotypical beliefs (as measured by the IAT) and emotional or behavioral responses, because attitudinal dissimilarity produces repulsion (Rosenbaum, 1986). Almost half a century ago, during the height of the civil rights movement, Rokeach and colleagues (e.g., Rokeach & Mezei, 1966) argued that racial prejudice was due more to perceived attitude dissimilarity between the races than differences in skin color. Studies that experimentally manipulated belief similarity and race found that perceived attitude dissimilarity produced biasing effects as strong or stronger than race (see Haidt, Rosenberg, & Horn, 2003; Insko, Nacoste, & Moe, 1983; Mezei, 1971). Moreover, the degree of Whites’ prejudicial attitudes varied according to the degree to which they perceived a dissimilarity of attitudes with African Americans (see Mezei, 1971).

Further, as Arkes and Tetlock (this issue) point out, attitudes and behavior termed *prejudice* may reflect reliance on culturally salient stereotypes that to some extent reflect social reality. For example, young African-American males are greatly overrepresented in the juvenile and criminal justice systems, a well-known statistical fact. (Although the reasons appear to be unrelated to race per se, see Redding & Arrigo, in press.) Unfortunately, “for many Americans, crime has a black face” (Armour, 1994, p. 787). Although most African-American males do not commit crimes, use of such base-rate information (often inherent in stereotypes) contributes to predictive accuracy. For prejudice researchers, however, use of base rate data linked to race is evidence of prejudice (Arkes & Tetlock, this issue). “It is difficult, even dangerous, to talk about group differences. ... Such candor is bound to provoke accusations of insensitivity and even racism. ... To evade these accusations and to be politically correct, social scientists avoid a frank dis-

cussion of significant cultural differences” (Ottati & Lee, 1995, p. 48). But in recent years, social psychologists have begun to investigate not just the process of stereotyping but the degree to which stereotypes are accurate (see Lee, Jussim, & McCauley, 1995). This recent research has shown that stereotypes can be accurate, that they do not typically exaggerate group differences, and that the use of stereotypes does not necessarily produce inaccurate judgments about individuals (McCauley, Jussim, & Lee, 1995).

Some conservative commentators (e.g., D’Souza, 1995) as well as rational-choice theorists (e.g., Farmer & Terrell, 2001) have suggested that attributions often characterized as racist may reflect a rational decision making calculus based on an intuitive actuarial risk assessment. D’Souza (1995) argued, for example, that the taxicab driver who declines to pick up an African-American man is not necessarily racist, but instead may be acting on his intuitive sense that young African-American males are more likely to commit crimes than others in the general population. (The statistical probability of being robbed by an African American is 7.5 times greater than the risk of being robbed by a White; see Arkes and Tetlock, this issue.) “How hollow it sounds to accuse cabdrivers of ‘prejudices’ and ‘stereotypes’ when their perceptions seem to be based on empirical reality” (D’Souza, 1995, p. 252). Khan and Lambert (2001) presented the taxicab example to White participants, who were asked to judge the rationality of the driver’s decision. Participants’ need for cognition (the tendency to engage in analytical thought) was assessed, and some participants were experimentally induced to think analytically about the scenario. Participants with a high need for cognition who were induced to think analytically were those most likely to judge the driver’s decision as rational. Thus, analytical thinking led to greater use of the base-rate information on differential crime rates across racial groups—in effect, to greater reliance on the racial stereotype and the base-rate information inherent in it.

It is surprising that so few researchers have previously raised or empirically examined the kind of compelling, yet facially obvious, interpretive critique of the implicit prejudice research provided by Arkes and Tetlock (this issue). Indeed, the failure to do so may reflect researchers’ own sociopolitical attitudes and biases. Research on implicit prejudice may be another example of liberal bias in psychological research; how one interprets research in this area may depend on whether he or she sits on the liberal or conservative side of the playing field. Hastorf and Cantril’s (1954) famous study of students’ perceptions of a Dartmouth versus Princeton football game found that Dartmouth students had very different perceptions of the same game than did Princeton students. It is much the same with research on an issue as politically and emotionally

charged as racial prejudice. Implicit prejudice researchers, like most psychologists who research social issues, have values invested in those issues (Redding, 2001). Liberal sociopolitical values likely influenced their problem definition and interpretation of research findings, both viewed through the pejorative lens of prejudice rather than through the more benign lens of stereotype knowledge. Psychologists have not devoted the same attention to testing and validating the benign explanation as they have the pejorative one, perhaps because the later advances a political agenda they find more appealing. That agenda, of course, is to provide evidence that racial prejudice remains widespread, strong, and somewhat intractable, thus bolstering the argument for social policies like affirmative action, which many of us support.

Moreover, although the research corpus on prejudice is voluminous, there is very little research on misattributions of prejudice. Sometimes referred to as stigma vulnerability, misattributions arise when minorities make attributions of prejudice against their group as an explanation for negative interpersonal outcomes in ambiguous situations (Gilbert, 1998, p. 306). Using the Prejudice Perception Assessment Scale, Gilbert found that African-American college students often made erroneous attributions of prejudice based on negative feedback they received in ambiguous social situations. Likewise, Inman and Baron (1996) found that African Americans were more likely to label an ambiguous act as reflecting racial prejudice and more likely to infer prejudice when the act was committed by a White toward a Black than vice-versa. Other studies (e.g., Rodin, Price, Bryson, & Sanchez, 1990) have produced similar results, including the finding that the actor–observer bias is more readily activated when the alleged perpetrator is White, to whom African Americans attributed great personal responsibility for the behavior (Flournoy, Prentice-Dunn, & Klinger, 2002). Taken together, these findings suggest that some African Americans may tend to stereotype Whites as prejudiced and that “those perpetrators who are expected to have prejudiced motives will be perceived to have prejudiced motives when none may exist” (Inman & Baron, 1996). To be sure, African Americans’ heightened sensitivity to racial prejudice is entirely understandable given this country’s long and sad history of racial injustice, but misattributions of prejudice are as corrosive for race relations as are prejudicial attitudes themselves. They also have negative effects on the individual perceiving the prejudice by increasing stress while decreasing one’s sense of self-efficacy (see Gilbert, 1998). Yet perhaps due to “the political view that prejudice is a ‘White problem,’ and to eliminate it, the mistaken attitudes of the White population must be changed” (Flournoy et al., 2002, p. 406), African-Americans’ stereotypes of Whites or misattributions of prejudice are rarely examined.

Arkes and Tetlocks's (this issue) critique is courageous, given the "politically sensitive" nature of research and public discourse on the emotionally charged issue of racial prejudice. Those who may be seen as challenging the view that racial prejudice remains endemic and widespread in the American culture and psyche run the real risk of being labeled racist, or at the very least, racially insensitive. They also may be seen as advancing a politically conservative research or policy agenda, something often not entirely welcome in the politically liberal social sciences (Redding, 2001). But as Arkes and Tetlock so persuasively argue, prejudices are distinct from stereotypes, which themselves may be partly anchored in social reality. To be sure, stereotypes can be so strong that they functionally operate as prejudice, particularly when they influence judgment and behavior, and stereotypes can be used all too readily to justify racial prejudice (Sniderman et al., 1998). As Stangor (1995) pointed out, "The misuse of stereotypes can have grave consequences for the victims of the stereotyping; thus, it behooves every one of us to think twice or even three times before using category membership as a basis of thinking about others" (p. 289). Individuals who are clearly nonprejudiced in the sense that they bear no ill will or discriminatory attitudes toward the stereotyped group still may act on the stereotypes they hold, producing the same result as if they had acted out of prejudice. Experimental and survey research has consistently found robust relations between negative stereotypes of African Americans and voters' policy preferences on social issues like affirmative action, crime control, and welfare policy (Carmines & Layman, 1998; Peffley & Hurwitz, 1998).

Many times a stereotype does equal a prejudice, but implicit prejudice research has not teased out the distinction, a vital task when considering the societal implications of understanding the prevalence and operation of racial prejudice in American society. Instead, implicit prejudice researchers equate stereotype knowledge with racial prejudice, perhaps reflecting their own sociopolitical biases on race relations in modern America, one of the most important but divisive issues of our time.

### Note

Richard E. Redding, Villanova University School of Law, 299 North Spring Mill Road, Villanova PA 19085-1597. E-mail: redding@law.villanova.edu

### References

- Armour, J. D. (1994). Race *ipsa loquitur*: Of reasonable racists, intelligent Bayesians, and involuntary negrophobes. *Stanford Law Review*, 46, 781-816.
- Carmines, E. G., & Layman, G. C. (1998). When prejudice matters: The impact of racial stereotypes on racial policy preferences of Democrats and Republicans. In J. Hurwitz & M. Peffley (Eds.), *Perception and prejudice: Race and politics in the United States* (pp. 100-134). New Haven, CT: Yale University Press.
- Dovidio, J. F., Mann, J., & Gaertner, S. L. (1989). Resistance to affirmative action: The implications of aversive racism. In F. A. Blanchard & F. J. Crosby (Eds.), *Affirmative action in perspective*. New York: Springer.
- D'Souza, D. (1995). *The end of racism: Principles for a multi-racial society*. New York: Free Press.
- Farmer, A., & Terrell, D. (2001). Crime versus justice: Is there a trade-off? *Journal of Law & Economics*, 44, 345-366.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality & Social Psychology*, 69, 1013-1027.
- Flournoy, J. M., Prentice-Dunn, S., & Klinger, M. R. (2002). The role of prototypical situations in the perceptions of prejudice of African Americans. *Journal of Applied Social Psychology*, 32, 406-423.
- Gilbert, D. J. (1998). The prejudice perception assessment scale: Measuring stigma vulnerability among African American students at predominately Euro-American universities. *Journal of Black Psychology*, 24, 305-321.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the implicit association test at age 3. *Zeitschrift für Experimentelle Psychologie*, 48, 85-93.
- Haidt, J., Rosenberg, E., & Horn, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, 33, 1-36.
- Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal & Social Psychology*, 49, 129-134.
- Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality & Social Psychology*, 70, 727-739.
- Insko, C. A., Nacoste, R. W., & Moe, J. L. (1983). Belief congruence and racial discrimination: Review of the evidence and critical evaluation. *European Journal of Social Psychology*, 13, 153-174.
- Judd, C. M., Park, B., Ryan, C. S., Brauer, M., & Kraus, S. (1995). Stereotypes and ethnocentrism: Diverging interethnic perceptions of African American and White American youth. *Journal of Personality & Social Psychology*, 3, 460-481.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality & Social Psychology*, 81, 774-788.
- Khan, S. R., & Lambert, A. J. (2001). Perceptions of rational discrimination: When do people attempt to justify race-based prejudice? *Basic & Applied Social Psychology*, 23, 43-53.
- Kinder, D. R. (1986). The continuing American dilemma: White resistance to racial change 40 years after Myrdal. *Journal of Social Issues*, 42, 151-171.
- Lee, Y., Jussim, L. J., & McCauley, C. R. (1995). *Stereotype accuracy: Toward appreciating group differences*. Washington, DC: American Psychological Association.
- McCauley, C. R., Jussim, L. J., & Lee, Y. (1995). Stereotype accuracy: Toward appreciating group differences. In Y. Lee, L. Jussim, & C. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 293-312). Washington, DC: American Psychological Association.
- Mezei, L. (1971). Perceived social pressure as an explanation of shifts in the relative influence of race and belief on prejudice across social interactions. *Journal of Personality & Social Psychology*, 19, 69-81.
- Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, 19, 395-417.
- Ottati, V., & Lee, Y. (1995). Accuracy: A neglected component of stereotype research. In Y. Lee, L. Jussim, & C. R. McCauley (Eds.),

- Stereotype accuracy: Toward appreciating group differences* (pp. 29–62). Washington, DC: American Psychological Association.
- Peffley, M., & Hurwitz, J. (1998). Whites stereotypes of Blacks: Sources and political consequences. In J. Hurwitz & M. Peffley (Eds.), *Perception and prejudice: Race and politics in the United States* (pp. 58–99). New Haven, CT: Yale University Press.
- Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tayler, R. B. (1990). Us and them: Social categorization and the process of ingroup bias. *Journal of Personality and Social Psychology*, 59, 475–486.
- Redding, R. E. (2001). Sociopolitical diversity: The case for pluralism. *American Psychologist*, 56, 205–215.
- Redding, R. E., & Arrigo, B. (in press). Multicultural perspectives on juvenile delinquency: Etiology and intervention. In C. Frisby & C. Reynolds (Eds.), *Handbook of multicultural school psychology*. New York: Wiley.
- Rodin, M., Price, J., Bryson, J., & Sanchez, F. (1990). Asymmetry in prejudice attribution. *Journal of Experimental Social Psychology*, 26, 481–504.
- Rokeach, M., & Mezei, L. (1966, January 14). Race and shared belief as factors in social choice. *Science*, 151, 167–172.
- Rosenbaum, M. E. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51, 1156–1166.
- Rudman, L.A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, 13, 79–82.
- Sniderman, P. M., Piazza, T., & Harvey, H. (1998). Prejudice and politics: An intellectual biography of a research project. In J. Hurwitz & M. Peffley (Eds.), *Perception and prejudice: Race and politics in the United States* (pp. 17–34). New Haven, CT: Yale University Press.
- Stangor, C. (1995). Content and application inaccuracy in social stereotyping. In Y. Lee, L. J. Jussim, & C.R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 275–292). Washington, DC: American Psychological Association.
- Tolerance.org. (n.d.) *Dig deeper: Test yourself for hidden bias*. Retrieved September 22, 2004, from [http://www.tolerance.org/hidden\\_bias/index.html](http://www.tolerance.org/hidden_bias/index.html)
- Zuriff, G. E. (2002). Inventing racism. *The Public Interest*, Winter, 114–128.

## A Perspective on Implicit Prejudice From Survey Research

David O. Sears

Department of Psychology  
University of California, Los Angeles

Arkes and Tetlock (this issue) seem animated more by a political than a scientific agenda. It criticizes the assertion that implicit prejudice reflects racial animus on the grounds that almost all college students today are reasonably fair-minded on racial issues, techniques of measuring implicit prejudice insure that nearly everyone will be classified as prejudiced and accused of harboring hidden biases for merely being aware of social realities, and decries the strongly moralistic tone sometimes used to condemn the implicitly prejudiced.

I come to my comments with three priors. First, at this point I am mainly a survey researcher and a user of conventional self-report measures of racial attitudes (i.e., explicit measures). The survey perspective seems to me important to get on the table when debating the measurement of racial attitudes. Second, I am also a social psychologist committed to empirical research. I happen to believe that the burden usually falls on a critic of an existing corpus of research to present some contrary evidence. Talk is cheap. Research is hard work, but it is social psychology's standard technique for establishing the value of our ideas. Third, my own research has been the target of quite similar critiques in the past (e.g., Sniderman & Tetlock, 1986; Tetlock, 1994). I considered recusing myself, but in the end felt that free debate is the best path to the truth.

I make three main points. First, Arkes and Tetlock's (this issue) article is entirely speculative. It presents no new research, cites no research done by the authors on

racial attitudes or prejudice, and proposes no new research. I offer some readily available evidence that contradicts some of its claims. Second, as mentioned, the article has much in common with earlier essays that aimed to discredit an earlier conceptualization and measurement of a new racism, asserting that it too was innocent of prejudice and unfairly stigmatized nonprejudiced people as racist. Those essays too were entirely speculative, but they led to considerable subsequent research. Much of it has been published outside the experimental social psychological literature, and I take this as a good venue for bringing it to the attention of experimentalists. Finally, I turn to the assumption in the implicit prejudice literature that explicit measures of racial attitudes are now too biased by social desirability pressures to be relied upon. Those pressures and biases may be real enough, but we need to take the next step, of evaluating whether or not they are of sufficient magnitude to jeopardize conclusions based on survey measures.

### Armchair Speculation

Arkes and Tetlock (this issue) most centrally asserts that implicit prejudice does not express racial animus, but may reflect shared cultural stereotypes, shame or guilt about racial inequality, or the savvy intuitive judgments of adept amateur statisticians. This sounds like the set up for a classic social psychological empiri-



cal question. But the style of argument is entirely speculative, dominated by *may* and *might* rather than *found* or *shown*. For example, negative emotions “could mean” anything from tragic to lazy, it is “easy to construct alternative explanations [of implicit prejudice results] ... one *might* be a tendency to be hostile ... or it might be shame or embarrassment,” “a White person who is genuinely ashamed of society’s treatment of African Americans by Whites *might well* be scored as prejudiced,” or that “Whites *may* nevertheless have a positive attitude toward African Americans, albeit not as positive as toward members of their own race” (emphases added). No new empirical evidence on these eminently testable ideas is presented.

Some relevant survey evidence from representative samples of adults is readily available, however. The first argument is that expressed stereotypes are cognitive—elements of social knowledge—rather than affective expressions of racial animosity. This common conceptual distinction seems logical enough in principle. One humbling lesson that survey research teaches us from time to time, however, is that not all distinctions compelling to social psychologists in theory are widely made by ordinary people in practice. In particular, they seem not to make a strong distinction between racial stereotypes and racial affect. First, in the 2000 National Election Studies (NES), scales of racial stereotypes (a difference score between ratings of Blacks and of Whites for intelligence, laziness, and trustworthiness) and of racial affect (the difference between ratings of Whites and Blacks on a 100-point feeling thermometer), are rather strongly correlated among Whites ( $r = .46$ ,  $n = 1139$ ). Second, both stereotypes and affect correlate at about the same level with symbolic racism, the most common and politically most powerful form of racial prejudice in America today ( $r = .38$  and  $.33$ , respectively), and with opposition to race-targeted policies, a key index of political impact ( $r = .22$  and  $.20$ , respectively, in the 2000 NES;  $r = .24$  and  $.27$  in the 1992 NES, per Sears, van Laar, Carrillo, & Kosterman, 1997). Third, factor analyses show that both stereotypes and racial affect load on a racial prejudice factor, distinct from (although correlated with) a conservatism factor composed of party identification and ideology (Sears & Henry, 2003). Speculations about a distinction between stereotypes and racial animosity are plausible enough, then. But as sometimes happens, the empirical findings do not cooperate; the best evidence is that the two dimensions operate similarly in the thinking of ordinary White adults.

A second assumption is that African Americans, such as those who took the Web-based IAT, cannot be prejudiced against their own race and are unlikely to endorse conventional unfavorable stereotypes about Blacks. However, survey data suggest that Blacks on average do tend to share the distinctive conventional stereotypes many Americans hold about Blacks, even

if not to the same derogatory degree (Bobo & Johnson, 2001; Sears, Citrin, Cheleden, & van Laar, 1999). Similarly, significant numbers of Blacks seem to show racial prejudice in racial affect. In the 2000 NES, 20% of the Blacks who gave different feeling thermometer ratings to Blacks and Whites rated Whites higher. Or, turn the question around: Are we safe in assuming without any further evidence that no Whites are prejudiced against their race? One might think that some might be critical of the White race’s history of ethnocentrism, racial insensitivity, and occasional imperial adventures against people of color. Indeed, of the Whites in the 2000 NES ( $n = 535$ ) who gave different feeling thermometer ratings to Blacks and Whites, 21% in fact rated Blacks more favorably.

Third, it is argued that individuals are fearful of being victimized by Black criminals because of rational base-rate information rather than because of their stereotypes and racial animosity. This does seem to ignore evidence that implicit prejudice correlates with such nonverbal behaviors toward Blacks as eye contact, eye blink, touching, and physical distance (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Fazio & Olson, 2003). Nevertheless, many Whites surely do appeal to base-rate information as the reason (or rationalization) for their discriminatory behavior: The legendary Los Angeles police chief of the 1960s, William F. Parker, frequently justified racially selective stop and search procedures as rational responses to known racial differences in crime rates. Are those rationales entirely free of racism?

Arkes and Tetlock’s (this issue) article draws an analogy to simple experiments showing that people often give too little weight to base-rate information (e.g., Kahneman & Tversky, 1973). The analogue seems to me lacking in ecological validity. Manipulations of base-rate information that consist of only a single cue, such as the proportion of lawyers or engineers, make for elegant and nicely controlled social psychological experiments, but for weak parallels with real-life decisions, which almost always involve multiple and often conflicting cues. Are police savvy intuitive statisticians if they use only a single demographic cue in deciding who to stop and search? Usually both superiors and the public hold them to more nuanced standards than stopping and searching for “driving while Black.” Even allowing the power of a single cue, where does White’s base-rate information about Blacks come from, and how accurate is it? The races remain highly segregated, so personal experience must not be much of a guide. The intense coverage on local television news of violent crimes committed by minorities seems to be an influential source of such information, although not of particularly accurate base rates (Gilliam & Iyengar, 2000). Similarly, the news media seem to have racialized welfare in a way that promotes quite in-

accurate base-rate perceptions of welfare recipients (Gilens, 1999).

Finally, Arkes and Tetlock (this issue) conclude that the overwhelming majority of White undergraduates score quite low on measures of prejudice, based on studies of students at two universities and using the Modern Racism scale developed in the late 1970s. A broader study, using an updated version of that scale, shows that both general population samples and students split fairly evenly on almost all items (Henry & Sears, 2002). At a more general level, Arkes and Tetlock imply that explicit measures of racial attitudes show that White Americans now harbor little racial animosity or resistance to racial equality, whereas implicit measures classify almost everyone as prejudiced. The former point is simply incorrect; some explicit measures have shown sharp decreases in racial animosity, but many others have not (Schuman, Steeh, Bobo, & Krysan, 1997).

### **A Familiar Theme**

Arkes and Tetlock's (this issue) article resembles an earlier effort by Sniderman and Tetlock (1986) in three ways. Both are critical of new lines of research describing an underrecognized but powerful form of racism in the post-civil rights era, both are wholly speculative, and both centrally claim that the new form of prejudice, as measured, has little to do with race and so unfairly stigmatizes the truly nonprejudiced. Much research has subsequently been done testing those earlier speculations. They turn out to have been largely inaccurate. Here I briefly summarize each charge and the relevant research.

Symbolic racism is not a coherent belief system in concept, measured with an incoherent set of items arbitrarily slapped together. Part of this was true, that it was originally inductively derived rather than tightly and deductively conceptualized. However in recent years it has consistently been conceptualized in terms of four themes: the denial of discrimination, criticism of Blacks' work ethic, and resentment of Blacks' demands and treatment by the broader society (Henry & Sears, 2002). Together they form a logically, psychologically, and statistically coherent belief system (Tarman & Sears, in press), measured with an updated scale including all four themes (Henry & Sears, 2002).

Symbolic racism is just old wine in new bottles, not materially different from old-fashioned or Jim Crow racism. That is incorrect by two standards. White support for old-fashioned racism (racial segregation, formal racial discrimination, and notions of biological inferiority) has sharply diminished, in most cases to the vanishing point (Schuman et al., 1997), whereas support for symbolic racism is quite widespread (Henry &

Sears, 2002; Tarman & Sears, in press). The political effects of symbolic racism dwarf those of old-fashioned racism and other forms of traditional prejudice (Hughes, 1997; Sears et al., 1997).

The theory that symbolic racism stems from anti-Black affect and conservative moral values (such as individualism) is unproven. That original theory has now been verified in several ways (Sears & Henry, 2003). A related claim was that symbolic racism reflects nonracial political conservatism rather than racial prejudice. To be sure, it is correlated with political conservatism. But symbolic racism loads about equally on racial prejudice and political conservatism factors in factor analyses including all three sets of variables (Sears & Henry, 2003). The best-fitting structural equation models require that symbolic racism be incorporated as a separate factor rather than allocating its items to other constructs like ideology (Tarman & Sears, in press). And controlling on ideology does not materially reduce the effects of symbolic racism on racially relevant dependent variables (Sears et al., 1997).

The associations of symbolic racism with racial policy preferences do not reflect its causal impact but are simply due to content overlap because both sets of items include references to government or affirmative action. Such associations in fact are nearly identical regardless of whether measures of symbolic racism include references to government (Tarman & Sears, in press). Measures of symbolic racism have not referred to affirmative action in over 20 years (see Sears & Citrin, 1982).

In other words, the earlier speculations almost all turned out to be incorrect once tested empirically (for reviews, see Hutchings & Valentino, 2004; Krysan, 2000; Sears & Henry, in press). The Arkes and Tetlock (this issue) article innocently cites that nearly 20-year-old predecessor as the definitive statement on these matters, but it is plainly badly dated. Whether the track record of this new article will prove similar is impossible to know. But they do share the same dubious core assumption that White Americans today are largely free of any significant levels of racial prejudice.

### **Social Desirability Biases in Explicit Measures of Racial Attitudes**

Research on implicit prejudice was partly stimulated by a methodological concern, that increasing normative pressures against expressing prejudice publicly had perhaps rendered conventional self-report measures of group-targeted prejudices unreliable (e.g., Fazio et al., 1995). If that were true, some widely cited survey-based findings might be jeopardized, such as the waning appeal of old-fashioned racism (Schuman

et al., 1997) or the continuing political impact of racism (e.g., Hughes, 1997; Kinder & Sanders, 1996; Sears et al., 1997). Here I would like to raise some cautions about prematurely rushing to that conclusion. Some such methodological biases have been reasonably well documented. But are they strong enough in a standard survey context to fundamentally threaten such findings, or are they just another source of error to be heeded?

The most obvious bias is that the truly racially prejudiced may overreport racially tolerant attitudes. Race-of-interviewer comparisons do find that White respondents generally express more positive racial attitudes to Black interviewers than to White interviewers, as if trying not to offend (Fazio et al., 1995; Kinder & Sanders, 1996; Schuman et al., 1997). Unobtrusive measures of racial attitudes, such as the bogus pipeline (Jones & Sigall, 1971) or the more recent list experiment (Kuklinski, Cobb, & Gilens, 1997), also have sometimes recorded more racial antagonism than do ordinary survey measures (also see Crosby, Bromley, & Saxe, 1980). Also the implicit prejudice literature itself suggested, at least initially, that true racism may be artificially suppressed on ordinary self-report measures, but leak out on implicit measures (e.g., Fazio et al., 1995).

Racial tolerance is no doubt overreported under some conditions. Is the bias of sufficient magnitude to explain important substantive findings? The race-of-interviewer studies show that Whites express more liberal attitudes to Black than to White interviewers. However, in most academic surveys White respondents rarely face Black interviewers. For example, in the 1986 NES survey, often analyzed in studies of racial attitudes, only 4% of the White respondents had non-White interviewers. At that low level any false racial liberalism elicited by Black interviewers could not bias the overall findings very much. Race of interviewer studies presumably are not informative about biases elicited by White interviewers. The list experiment has not been the subject of validation studies, but some of its estimates raise questions (e.g., that 98% of White Southerners expressed covert anger about affirmative action in the early 1990s; Kuklinski et al., 1997). Finally, implicit prejudice may not be so centrally relevant to political psychology, because it seems to be more closely linked to automatic nonverbal behaviors than to such deliberate and thoughtful behaviors as expressed political choices (Dovidio et al., 1997; Fazio et al., 1995). Also, studies relating implicit and explicit of racial prejudice have so far turned up quite mixed results (Fazio & Olson, 2003).

A second idea is that social desirability pressures may lead some Whites to express no opinion at all rather than expressing their true prejudices (Berinsky, 2004). Are ordinary survey findings substantially dis-

torted because large numbers of racial conservatives give noncommittal responses instead? The original research on this point mainly used items that draw unusually large numbers of nonresponse because they are preceded by a rare and draconian screening option intended to eliminate people without firm attitudes (e.g., have you had enough interest in this question enough to favor one side or the other?). In the 1986 NES, on items with such screeners, an average of 33% of the Whites chose the no-opinion option; on items with no such screeners, the average was 3% on the symbolic racism items, 3% on the old-fashioned racism items, and 6% on racial policy items. It seems that relatively little prejudice must go unreported in most surveys due to true racists avoiding any response.

A third possible bias might inflate the estimated impact of racial prejudice on racially relevant political preferences. Systematic individual differences in responsiveness to social desirability pressures could produce correlated under-reporting of racial animosity on both independent and dependent variables, artifactually inflating the estimated association between them. However, two kinds of evidence suggest that any such dynamic is not likely to be strong enough to explain the strongest political effects of racial prejudice. First, such correlated errors would inflate the predictive effects of all indicators of racial prejudice; a rising tide should lift all boats. But different indicators of prejudice yield associations with racial policy preferences of very different strengths. Symbolic racism consistently has far stronger effects than do old-fashioned racism, stereotypes, or racial affect (e.g., Hughes, 1997; Sears et al., 1997). Second, Black interviewers should trigger social desirability pressures more than White interviewers should. By this logic, Black interviewers should elicit artifactually higher associations between racial prejudice and racial policy preferences than White interviewers should. However, in the 1986 NES, the bivariate correlations between symbolic racism and racial policy preferences were  $r = .56$  for those interviewed by non-Whites as against  $r = .57$  for the full White subsample (Tarman & Sears, in press). It is possible, of course, that even White interviewers may induce some social desirability pressures. But surely the non-Whites should have induced more, and they seem not to have.

These are mere cautionary comments, not strong assertions. Social desirability biases can be shown in experiments, and no doubt do occur to some extent in surveys. Neither tells us whether they occur in *sufficient magnitude* to threaten the main findings of survey research. That is the bottom-line question about any source of error in psychological research. At this point I am skeptical and have offered some preliminary data to back that up. Much more research is needed. My primary suggestion is to avoid premature closure on issues that have not yet been thoroughly explored.

## Notes

I thank David Amodio, Vika Savalei, and Christopher Tarman for their generous help with the preparation of this article.

David O. Sears, Department of Psychology, 1285 Franz Hall, UCLA, Box 951563, Los Angeles, CA 90095-1563. E-mail: sears@issr.ucla.edu

## References

- Berinsky, A. J. (2004). *Silent voices: Public opinion and political participation in America*. Princeton, NJ: Princeton University Press.
- Bobo, L., & Johnson, D. (2001). Racial attitudes in a prismatic metropolis: Mapping identity, stereotypes, competition, and views on affirmative action. In L. D. Bobo, M. L. Oliver, J. H. Johnson, Jr., & A. Valenzuela, Jr. (Eds.), *Prismatic metropolis: Inequality in Los Angeles* (pp. 81–163). New York: Russell Sage Foundation.
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice. *Psychological Bulletin*, 87, 546–563.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33, 510–540.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.
- Gilens, M. (1999). *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. Chicago: University of Chicago Press.
- Gilliam, F. D., Jr., & Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science*, 44, 560–573.
- Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23, 253–283.
- Hughes, M. (1997). Symbolic racism, old-fashioned racism, and Whites opposition to affirmative action. In S. A. Tuch & J. K. Martin (Eds.), *Racial attitudes in the 1990s: Continuity and change* (pp. 45–75). Westport, CT: Praeger.
- Hutchings, V. L., & Valentino, N. A. (2004). The centrality of race in American politics. *Annual Review of Political Science*, 7, 383–408.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349–364.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kinder, D. R., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*. Chicago: University of Chicago Press.
- Krysan, M. (2000). Prejudice, politics, and public opinion: Understanding the sources of racial policy attitudes. *Annual Review of Sociology*, 26, 135–68.
- Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the “New South.” *Journal of Politics*, 59, 323–349.
- Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations* (Rev. ed.). Cambridge, MA: Harvard University Press.
- Sears, D. O., & Citrin, J. (1982). *Tax revolt: Something for nothing in California*. Cambridge, MA: Harvard University Press.
- Sears, D. O., Citrin, J., Cheloden, S. V., & Van Laar, C. (1999). Cultural diversity and multicultural politics: Is ethnic Balkanization psychologically inevitable? In D. Prentice & D. Miller (Eds.), *Cultural divides: The social psychology of cultural contact*. New York: Russell Sage Foundation.
- Sears, D. O., & Henry, P. J. (2003). The origins of symbolic racism. *Journal of Personality and Social Psychology*, 85, 259–275.
- Sears, D. O., & Henry, P. J. (in press). Over thirty years later: A contemporary look at symbolic racism. In M. Zanna (Ed.), *Advances in experimental social psychology*.
- Sears, D. O., van Laar, C., Carrillo, M., & Kosterman, R. (1997). Is it really racism? The origins of White Americans’ opposition to race-targeted policies. *Public Opinion Quarterly*, 61, 16–53.
- Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political debate. *Journal of Social Issues*, 42, 129–150.
- Tarman, C., & Sears, D. O. (in press). The conceptualization and measurement of symbolic racism. *Journal of Politics*.
- Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology*, 15, 509–529.



## Racism in the Brain; or Is It Racism *on* the Brain?

Peter Suedfeld

*Department of Psychology  
The University of British Columbia*

My first reaction after reading the target article (Arkes & Tetlock, this issue) was that my comment would be brief indeed: They're right. However, the need to write such an article intrigued me more. Exactly a decade ago, Tetlock (1994) published a telling criticism of a previous attempt to infer racism from data that had little if any face validity as measures of that construct; now, he is taking part in criticizing another generation of such attempts. Each of these approaches has generated scores of studies and no doubt will continue to do so regardless of what flaws are pointed out in the theory, in the methodology, or in the inferences drawn from it. Why?

The study of racism has a long history in American social psychology, personality, and political psychology. One of the fascinating aspects of this massive, and ever-growing, body of research is the repeated need for new and supposedly more subtle measures of prejudice as existing ones fail to support the view that American Whites are irredeemably racist. This is no place for a full review of the literature, so I shall just mention a few well-known examples. Because the target article, and its targets, focus on attitudes toward African Americans, I direct this overview to the same topic, and I use the ethnic terms in vogue at the time of the original publications. Let us look briefly at various measures of racism in the social psychological literature.

### Ethnic Stereotypes

Measures of stereotyped attitudes toward minority groups began when Katz and Braly (1933) asked Princeton University students to indicate what traits are considered to be typical of various ethnic and national groups. Generally positive traits were ascribed to Americans, Europeans, and Japanese, and generally negative ones to Negroes, Jews, Turks, and Chinese.

Replications in subsequent years showed diminishing consensus about stereotype content (Gilbert, 1951); stereotypes became more nuanced—for example, they were seen to differ by sex within ethnic groups—and, on the whole, stereotypes showed fewer negative and more positive traits except for groups with which the United States was then in conflict. Some respondents commented that the whole idea of stereotyping ethnic groups was unreasonable (Karlins, Coffman, & Walters, 1969). The stereotype of Blacks became successively more positive (Dovidio & Gaertner, 1986; Karlins et al., 1969).

A more recent assessment (Devine & Elliot, 1995) interpreted the research as indicating the respondents' personal beliefs rather than their understanding of societal stereotypes, implying that prejudice decreased during these five decades. Furthermore, Devine and Elliot reported that one-fifth of their subjects refused to fill out a form asking explicitly about their own beliefs concerning Blacks, on the basis that one cannot characterize a whole group. Thus, there may still be a negative stereotype of Blacks in America, but it seems that personal beliefs in this stereotype have waned substantially.

### Social Distance

Published in the same year as the Katz and Braly (1933) study, the Bogardus (1933) Social Distance Scale provided a hierarchy of social contacts (from marriage to living in the same country) that respondents would accept vis-à-vis other ethnic groups. In a study predating the full presentation of the scale, Bogardus (1925) reported on the social distance tolerances of a group of White schoolteachers and business people. On Bogardus's 7-point ordinal scale, Negroes were rated at an average score of 5.10, between allowing them to be citizens of the country and only allowing them as visitors.

Some 70 years later, Kleg and Yamamoto (1998) found the mean rating for African Americans to be 1.55, between acceptance as a best friend and "to marry into the group." Although the ranking of African Americans among the groups had not changed much, their acceptability had obviously made dramatic gains. This was consistent with generally greater social acceptance of other groups and less differentiation among them. One could argue that the persistence of a hierarchy shows lingering prejudices (cf. Sidanius & Pratto, 1993), but the much flatter hierarchy shows a great reduction in prejudice.

### Direct Attitude Measurement

*The Authoritarian Personality* (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) is a seminal work, one that every psychology student learns about and every reviewer cites. The original work included two measures of racism: the Anti-Semitism (A-S) Scale and the Ethnocentrism (E) Scale. The A-S Scale had subscales measuring various negative

feelings about Jews; the E Scale had subscales for attitudes toward Negroes, toward minorities in general, and patriotism. Some of the items are unrelated to ethnic prejudice (e.g., support for the abolition of minor political parties, support for hierarchical relations among nations, hostility toward conscientious objectors and draft evaders).

Neither scale has been used much in recent times. Both the original version of their replacement, the F (for Fascism) Scale, and its later variants are still popular. Adorno et al.'s (1950) F Scale has no questions directed at prejudice against specific ethnic groups (there is one question referring to "Germans and Japs" in the context of what to do after end of World War II). This is reasonable enough: The F Scale is supposed to be a measure of authoritarianism (i.e., susceptibility to antidemocratic political movements). The authors (unlike some later scholars) recognized that ethnocentrism can comfortably coexist with the politics of the Left as well as the Right (Adorno et al., 1950, pp. 189–190). Empirically, the F Scale correlated positively and moderately with the A-S Scale ( $r = 0.53$ ) and—in its final version—somewhat more highly with the E Scale ( $r = 0.73$ ).

Critics have savaged the F Scale's methodological and psychometric shortcomings, its susceptibility to response biases unrelated to authoritarianism, and its excessive focus on authoritarianism of the Right while ignoring that of the Left (e.g., Eysenck, 1954). Adorno et al. (1950) were less guilty of such a blinkered view than they were accused of being, although their nuanced perspective was indeed ignored by some other users of the F Scale. At any rate, the critiques were themselves criticized, and the controversy spawned revisions of the F Scale (Hyman & Sheastley, 1954; Rokeach, 1956) that tried to balance both the question format (to eliminate response bias) and question content (to eliminate political bias).

Clearly, the F Scale is not a measure of prejudice, but even its correlation with prejudice is arguable. For example, a recent report indicated that the scale showed no significant relation with anti-Semitism. It, or its successors, may be useful measures of orientation toward authority and social conventions, but it is the nature of authority and society that should then dictate the attitudinal content of highly authoritarian individuals.

One example is the currently most popular revisionist work, Altemeyer's (e.g., 1981) Right-Wing Authoritarianism (RWA) Scale, which has serious conceptual and methodological problems of its own (see Martin, 2001). Its defining components parallel some foci of the original authoritarian personality construct: adherence to the dictates of authority and to societal conventions and hostility toward those who fail to so adhere. A close relation between authoritarianism and prejudice would be expected if the outgroup is perceived as

violating or threatening the conventional norms of society (Feldman, 2003) or if society or its leaders endorse, reinforce, and accept racial prejudice. By the same token, high RWA scorers should be low in prejudice when the other group (or groups) are seen as respecting authority and upholding society's norms and when society as a whole and its leaders reject, suppress, and punish prejudicial attitudes.

Most American (and other Western) leaders and societies in the past half-century or so have in fact taken this route, and we would expect that prejudice would have decreased even among people who score high on the authoritarianism measures. In fact, just such people should be showing the largest reductions in ethnic and racial prejudice over that period. Although there may be no longitudinal test of that specific hypothesis, scores on straightforward self-reports of racist attitudes have indisputably decreased during the past several decades (for a brief review, see Dovidio & Gaertner, 1996). The attempt to show high levels of racism among White Americans had to abandon the use of direct attitude measures.

### Indirect Measures of Racism

Some social psychologists have accepted the evidence that direct measures of prejudice have shown substantial decreases (cf. Roth, 1994), but remain skeptical about whether these changes reflect an actual reduction in prejudicial attitudes. They have developed new concepts of racism and new tests to measure these.

One early contribution to this effort was the invention of *modern* or *symbolic* racism (McConahay, 1986; Kinder & Sears, 1981). Sears (1997) has contrasted this attitude, "hostility toward or discrimination against blacks," with what he calls "'old-fashioned' or 'redneck' or 'blatant' racism": the belief that race determines one's traits, and that some races are biologically superior to others (pp. 217–218). According to Sears, the great majority of American Whites have retained the former attitude while repudiating the latter.

This argument, and the measures used by its advocates (e.g., the Modern Racism Scale; McConahay, 1986), have attracted considerable criticism (Arkes & Tetlock, this issue; Martin, 2001; Roth, 1994; Tetlock, 1994). Without repeating these in detail, and ignoring purely methodological issues, we can see that these scales share the problem posed by using F and RWA as measures of racism. Specifically, they infer the existence of racism from attitudes that may at least as reasonably be seen as emanating from belief-opinion systems whose causal relation to prejudice is arguable. Political conservatism is one such system; others are source beliefs fundamental in American society: individualism, self-reliance, achievement orientation, meritocratic values, and a strong work ethic. Specific

attitudes growing out of these sources might include opposition to race-based affirmative action and school busing, a rejection of historical group guilt, and the belief that poverty and crime are caused by personality and motivation not by discrimination.

The counterargument has been that traditionalism and racism make separate contributions to these attitudes, with the latter being preponderant (e.g., Sears, 1994). Nevertheless, as Tetlock (1994) pointed out, it seems that racism *per se* is mostly still measured by items tapping “old-fashioned” racism. Combining the results with tests of traditional views confuses the issue—although it does make it possible to malign conservatives by association. The correlation between racial prejudice and these indirect measures (and even such scales as F and RWA) may be confounded by the perception of some Whites (and not only Whites; cf. the Jesse Jackson anecdote in the target article) that Blacks are statistically more prone to violating traditional values. Actual racism might be inferred if the respondent is more negative toward Black than toward White violators of these norms, and the evidence on that is very mixed.

### Measures of Latent Racism

The attempt to infer racism from measures totally unrelated to the expression of attitudes is another step in distancing the conclusions from the data. There are two major problems with such techniques: One is the level to which they are valid measures of racism, and the other is whether they have any real-life (as opposed to statistical) significance.

The Arkes and Tetlock (this issue) critique addresses two cognitive measures of latent racism: affective priming and implicit association (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998). Dovidio and Gaertner (1996), in reviewing such studies, labeled differences in reaction time and in nonverbal behaviors, such as differential eye contact, unintentional bias; perhaps they are, but—as Arkes and Tetlock discuss in detail—they may also be due to other sources. Response latency has been used as an indicator of attitudes in a variety of contexts, including telephone surveys of political opinions, and a number of possible confounds and alternative explanations have emerged (Bassili, 2000).

Other latent indexes have their own problems. For example, bias has been inferred from finding that Whites make more direct eye contact when speaking with other Whites than when speaking with Blacks. Aside from bias, this may be due to embarrassment and other emotions. It may also exemplify reciprocity: African Americans (as well as some Asian and African groups) tend to avoid direct eye contact during conversation (Axtell, 1991). The proposition that Whites who avoid eye contact more do so because of racism is no

more inherently logical than that they are just more socially sensitive to cues from other person. Another approach is to infer latent racism from physiological (including neurophysiological) processes. Among the most up-to-date attempts is the use of brain imaging (Phelps & Thomas, 2003). Aside from facial recognition studies, the most relevant of these may be Phelps and Thomas’s showing pictures of Black and White faces and using neuroimaging to measure activation in the amygdala, a learned fear response that may not be accompanied by any other fear reaction. Results of studies using this approach have been inconsistent, and the neural response is affected by the familiarity or unfamiliarity of the stimulus (as is facial recognition, of course).

One interesting example of this technique is a study by Phelps et al. (2000). Besides recording the activity of the amygdala, the authors correlated that response with other measures of racism: the Modern Racism Scale (McConahay, 1986), the Implicit Association Test (Greenwald et al., 1998), and a measure of the startle reflex to Black or White faces (which is also affected by the familiarity of the stimulus). Thus, Phelps et al. attempted to validate a new, neurological measure against three older measures, all of which are themselves of arguable validity and susceptibility to artifacts.

The results were intriguing, and somewhat amusing. The Modern Racism Scale showed pro-Black bias, the IAT showed anti-Black bias, and neither the startle test nor brain activation showed a significant difference in either direction. The latter did, however, correlate with IAT and startle scores, although not with the Modern Racism Scale (which, strangely, Phelps and Thomas, 2003, characterize as an explicit test of racism—and which it is explicitly not supposed to be). Phelps and Thomas, with commendable restraint (not shown by all brain researchers) point out that “it is inappropriate to assume that the results of neuroimaging studies of a given behavior are more informative than the results of psychological studies of that behavior” (p. 747).

### Putting the Problem in Context

It seems obvious that American society has made major strides to reduce discrimination against Blacks within the past half-century or so. That period has seen the end of legally mandated discrimination—segregation, debarment from many public and private facilities (universities, restaurants, clubs, beaches, golf courses, etc.), military assignments mostly to menial work in noncombat units, obstacles to voting, laws against miscegenation, and so on—and its replacement by legally mandated nondiscrimination. In fact, the currently lively arguments are about the extent to which institutions including government should favor African

Americans in such areas as employment, higher education, and government contracts. One might argue that such institutional changes have eliminated neither the disadvantages that Blacks face nor the prejudiced attitudes of individual Whites, but the widespread advocacy of these changes by eminent Americans in all walks of life, and their acceptance by society at large (e.g., the success of African Americans at gaining elective office in racially mixed areas) are strong indications that attitudes have in fact undergone drastic alteration.

This observation concerning differences across time may be supplemented by looking at differences across geography. In a world that has seen, comparatively recently, the massacres of Armenians by the Turks, Jews by the citizens of most European nations, Tibetans by the Chinese, Tutsis by Hutus, Bosnians by Serbs and vice versa, and— at lower numbers, but equally lethally—ethnic or religious murders all over the world, it seems strange to obsess about racism as revealed only by slower reaction times or brain activation.

The fact that things are much worse elsewhere, and have been much worse here, would not justify the persistence of discrimination. However, one must question whether even if real, the subtle, symbolic, and latent forms of racial dislike are of sufficient importance to call for so much attention from social scientists. One might well ask why our disciplines have these remnants of the old troubles so much in mind (or on the brain). Is it that we can't get away from Gunnar Myrdal's (1944) identification of anti-Negro prejudice as the central dilemma of American society, or do social scientists by nature focus on negative aspects of life, even as those negative aspects diminish? It is perhaps time to study the "positive psychology" converse of all this: How it is that White Americans, many of whom can personally remember the days of Jim Crow and all of the abuses of racial equality, have managed to put most of those practices and actions behind them and—unlike citizens of so many other countries—dramatically reduced, if not eliminated, one of the long-lasting and pervasive evils of their society.

### Note

Peter Suedfeld, Department of Psychology, University of British Columbia, 3533 – 2136 West Mall, Vancouver, B.C. V6T 1Z4. E-mail: psuedfeld@psych.ubc.ca

### References

- Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper & Row.
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg: University of Manitoba Press.
- Axtell, R. E. (1991). *Gestures: The do's and taboos of body language around the world*. New York: Wiley.
- Bassili, J. N. (Ed.). (2000). Response latency measurement in telephone surveys [Special issue]. *Political Psychology*, 21, 1–132.
- Bogardus, E. S. (1925). Measuring social distance. *Journal of Applied Sociology*, 9, 299–308.
- Bogardus, E. S. (1933). A social distance scale. *Sociology and Social Research*, 17, 265–271.
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton Trilogy revisited. *Personality and Social Psychology Bulletin*, 21, 1139–1150.
- Dovidio, J. F., & Gaertner, S. L. (1986). Prejudice, discrimination, and racism: Historical trends and contemporary approaches. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 1–34). New York: Academic.
- Dovidio, J. F., & Gaertner, S. L. (1996). Affirmative action, unintentional racial biases, and intergroup relations. *Journal of Social Issues*, 52, 51–75.
- Eysenck, H. J. (1954). *The psychology of politics*. London: Routledge.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Feldman, S. (2003). Enforcing social conformity: A theory of authoritarianism. *Political Psychology*, 24, 41–74.
- Gilbert, G. M. (1951). Stereotype persistence and change among college students. *Journal of Personality and Social Psychology*, 46, 245–254.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Hyman, H. H., & Sheatsley, P. B. (1954). "The authoritarian personality": A methodological critique. In R. Christie & M. Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality"* (pp. 50–122). Glencoe, IL: Free Press.
- Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 13, 1–16.
- Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280–290.
- Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40, 414–431.
- Kleg, M., & Yamamoto, K. (1998). As the world turns: Ethno-racial distances after 70 years. *Social Science Journal*, 35, 183–191.
- Martin, J. L. (2001). *The authoritarian personality*, 50 years later: What lessons are there for political psychology? *Political Psychology*, 22, 1–26.
- 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.
- Myrdal, G., with R. Sterner & A. Rose (1944). *An American dilemma: The Negro problem and modern democracy*. New York: Harper & Brothers.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. 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.
- Phelps, E. A., & Thomas, L. A. (2003). Race, behavior, and the brain: The role of neuroimaging in understanding complex behaviors. *Political Psychology*, 24, 747–758.
- Rokeach, M. (1956). Political and religious dogmatism: An alternative to the authoritarian personality. *Psychological Monographs*, 70(18).



- Roth, B. M. (1994). *Prescription for failure: Race relations in the age of social science*. New Brunswick, NJ: Transaction.
- Sears, D. O. (1994). Ideological bias in political psychology: The view from scientific hell. *Political Psychology*, 15, 547–556.
- Sears, D. O. (1997). Review of B. M. Roth (1997). *Prescription for failure: Race relations in the age of social science*. *Political Psychology*, 18, 209–215.
- Sidanius, J., & Pratto, F. (1993). The inevitability of oppression and the dynamics of social dominance. In P. Sniderman & P. E. Tetlock (Eds.), *Prejudice, politics, and the American dilemma* (pp. 173–211). Stanford, CA: Stanford University Press.
- Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good intentions? *Political Psychology*, 15, 509–529.

## Implicit Prejudice: Pentimento or Inquisition?

William von Hippel

School of Psychology  
University of New South Wales

A few years ago I attended a Person Memory Interest Group preconference at which the Implicit Association Test (IAT) featured prominently. At the end of the preconference, Brian Nosek and Wil Cunningham stopped to ask my opinion of the IAT. As they put it, “Most people are either for or against it, but we can’t decide where you stand.” This comment reveals a striking uniqueness about the IAT in recent social psychological history. Rarely has a methodological tool garnered such strong adherents and detractors. Scientists are often emotional advocates or critics of new theoretical approaches, but methods are typically less divisive. Why has the IAT led to such polarization of the research community?

With regard to positive feelings, I think the reasons for the excitement are clear. The IAT provides a clever and simple procedure for measuring implicit attitudes that typically generates huge effect sizes (more on this issue later). Indeed the IAT effect is so strong that it is unique among implicit measures in that naïve participants can discern its purpose by virtue of the difficulty they experience during the “incompatible” trials (more on this later too). Like many social psychologists, I use a version of the IAT in which students tap on their desks as a classroom demonstration of implicit ageism, and nervous laughter erupts every year when students first attempt the incompatible trials and are suddenly and audibly slowed down. For these reasons, many of us are excited about the potential of the IAT to expand the study of implicit social cognition. We still do not really understand what it reveals and why it works (although Riki Conrey, Jeff Sherman and their colleagues, 2004, have some promising models that separate its component processes), but the IAT may yet prove to be one of the most important methodological advances in social cognition.

Negative attitudes toward the IAT are also pervasive, however, and Arkes and Tetlock (this issue) have clearly articulated why. To label an unconscious response, and especially one that is still poorly under-

stood, as *prejudice* strikes many as irresponsible (in this regard, I too am guilty, as my colleagues and I have often used that label; e.g., von Hippel, Sekaquaptewa, Vargas, 1997). The use of this label also suggests that political views are leaking into psychological research, as the agenda served by implicit prejudice research is largely a liberal one. So, with the proviso that I believe that Arkes and Tetlock have done the field a great service by opening debate on the meaning of *implicit prejudice*, let me turn now to the specifics of their argument. I focus on areas in their article where I think they have stretched an argument too far and on areas where I think they have not gone far enough.

### Arkes and Tetlock are Too Bold ...

#### Rationality of Prejudice

In their discussion of expected utility theory, Arkes and Tetlock suggest that prejudice and discrimination can be rational if societal-level data suggest that members of different groups are likely to behave in different ways. They provide a compelling example of the nervous Rev. Jackson and mathematically derive that he should be much more likely to flee when followed by a Black than a White pedestrian at night. But their example is missing two ingredients that are common in real-life applications of this principle: individuating information, which can be much more diagnostic than category-level information, and cost to the target.

Regarding individuating information, let me relate an incident involving one of my students in a course on prejudice. After class one day she went to withdraw cash, but became nervous when a group of young Black men was hanging around the ATM. Remembering my exhortations from lecture, she decided to be “unbiased” and withdrew her money, at which point she was robbed. I asked her to describe the men who were hanging out by the ATM, and then asked her if she would have withdrawn the money if they looked

and behaved identically but were White. She said she would not, so I suggested that these individuating cues were the relevant data that should have influenced her behavior. The fact that the men had Black skin was far less predictive of their probability of taking her money than the fact that they dressed and acted like thugs. Bending over backward to be nonprejudiced in this circumstance was clearly a poor idea, but only because she relied on category-level data rather than the more diagnostic individuating cues that were available.

We also need to consider more fully the cost to targets of discriminatory behavior. There is virtually no cost in the nervous pedestrian example, but disutility rises when targets suffer as Tetlock once pointed out to me in a discussion of liberty versus responsibility (i.e., my right to swing my fist ends at your chin). For this reason, the rationality of discrimination is far less clear in the case of the nervous employer, the nervous landlord, or the nervous banker. Even if we place no utility on egalitarian concerns, rationality suggests that failure to employ or house a significant percentage of the population will create a self-fulfilling prophecy that will be costly to all members of society.

### Discounting

Arkes and Tetlock (this issue) point out that it is irrational not to discount ability attributions to recipients of affirmative action. They raise an excellent point, but they fail to note that affirmative action is a psychological sword that cuts only one way. Many members of society are beneficiaries of affirmative action, but only members of stigmatized groups suffer ability discounting as a consequence. Recipients of affirmative action who are not chronically stereotyped as inferior are generally unbothered when they receive affirmative action and their abilities are not discounted by others (see Crocker, Major, & Steele, 1998). Thus, I would agree that discounting of minority affirmative action recipients should be at least partially offset by augmenting for obstacles overcome, whereas discounting without augmentation should be applied in corporate cases of hiring personal or family connections, and university cases of legacy admits, student athletes, and all the other beneficiaries of preferential admission and hiring practices who are not visibly identifiable by virtue of their membership in chronically stigmatized groups. The fact that stigmatized recipients of affirmative action themselves tend to discount other ingroup recipients is greater testimony to the power of stigmatization than it is evidence of appropriate discounting.

### Problem of Relativity

Arkes and Tetlock (this issue) suggest that we ought not use the label of *prejudice* if we cannot distinguish on the IAT whether performance reflects liking of both

groups but preference for one's own group versus liking of one's own group and disliking of the outgroup. Although disliking an outgroup is more clearly prejudicial than simply liking one's own group more, there is a long tradition of treating relative preferences as prejudicial. Allport (1954) derived from Spinoza the suggestion that "love prejudice" (ingroup positivity) is not only more prevalent than "hate prejudice" (outgroup negativity), but is also the source from which hate prejudice springs. More recently a number of social psychologists have suggested that various types of discrimination previously thought to arise from outgroup negativity may instead be a function of relative ingroup positivity (Brewer, 2001). Thus, the inability of the IAT to distinguish between outgroup negativity and ingroup positivity limits its utility, but does not mean that it is not tapping prejudice. Additionally, it should be noted that Karpinski and Steinman (2004) have adapted the IAT into the Single Category Association Test, which appears to have promise for separating ingroup and outgroup components of implicit prejudice.

### Arkes and Tetlock are Too Timid ...

#### Measures    Constructs

To study a construct it is, of course, necessary to measure it. But intense focus on a measure carries with it the risk of unintentionally conflating the measure with the construct. A great deal of evidence suggests that implicit attitudes do exist, but that does not mean that factors that influence a particular measure of implicit attitudes necessarily influence the attitude itself. Researchers are well aware of this distinction in the case of explicit attitudes, as few would argue that an explicit attitude has been changed if people report different attitudes when under duress. Nevertheless, many accept that an implicit attitude has changed if people show movement on the IAT. For example, consider the following sets of experiments: (a) Dasgupta and Greenwald (2001) found that exposure to admired Blacks and disliked Whites led to a reduction in the typical race IAT effect for up to 24 hr; (b) Lowery, Hardin, and Sinclair (2001) demonstrated reduced implicit prejudice when White participants were tested by a Black rather than a White experimenter; and (c) Blair and Lenton (2001) demonstrated that implicit stereotyping is reduced when people imagine counter-stereotypic targets prior to the implicit stereotyping measurement. Blair and Lenton's research is noteworthy in that they measured implicit stereotyping via the IAT, the GNAT, and the Deese/Roediger-McDermott false memory paradigm, with similar results across all three measures.

Although these findings are provocative and incredibly interesting, it is unclear what they mean. They show

that the IAT and other implicit measures are malleable to manipulations of accessibility, but whether that indicates that implicit attitudes are similarly malleable is a separate question. My bet is that the intuition many of us had when this work began—that implicit attitudes are much harder to change than explicit attitudes—may indeed be true, and that these data may not directly implicate change in implicit attitudes themselves. Rather, these data may demonstrate that implicit attitude measures are as easy to move around as explicit ones (although different procedures are required) and that movement on the measure does not necessarily indicate movement in the underlying attitude.

### **An Effect That Is Too Big to be True, Probably Isn't**

Why do most Whites and Blacks show anti-Black bias on the IAT? Arkes and Tetlock (this issue) expand on Karpinski and Hilton's (2001) suggestion that the race IAT taps familiarity with cultural stereotypes, rather than an implicit attitude toward Blacks. An alternative, although not mutually exclusive, perspective can be found in a recent article by Kinoshita and Peek-O'Leary (2004). Kinoshita and Peek-O'Leary suggested that compatibility effects in the IAT between *pleasant* and a target category such as *white* could arise in part from the default nature of the category *white* relative to the contrasting category *black*, rather than reflecting a conceptual association between the target category and pleasantness. In support of such an account, they replicate the *insect/flower* IAT effect (albeit with a reduced effect size) when the intervening *pleasant/unpleasant* judgment is replaced with a *word/not word* judgment.

These data suggest that caution should be exercised in assuming that a race IAT effect that is greater than zero reflects differential implicit preference for Whites versus Blacks, as figure-ground asymmetries appear to produce a reliable IAT effect when no evaluation is involved. An important advantage of the IAT was that it seemed to be a ratio scale, with a true zero value, but it may be the case that recalibration is necessary to recover this true zero value. In the meantime, the finding that a majority of respondents show race bias in the IAT is open to alternative interpretation.

The location of the true zero value of the IAT is further clouded by research that suggests that the IAT can itself induce stereotype threat in White students. Specifically, Frantz, Cuddy, Burnett, Ray, and Hart (in press) showed that White students find the IAT threatening if they are told that it is a measure of prejudice, or if they ascertain that themselves as they take the test. In this case, stereotype threat emerges from participants' concern that they will be perceived as prejudiced because they are White. This threat disrupts performance on the IAT, much as it does with other groups in other

performance domains, by causing people to provide even larger IAT effects than they would otherwise. Again, these findings call into question the interpretation of effect size and the location of the true zero value on the IAT.

### **A Shameless Plug for My Own Research**

In contrast to the dominant methods for studying implicit attitudes in social psychology, my approach to the study of implicit attitudes has its intellectual roots in the work of Roediger (1990), who suggested that the implicit/explicit memory distinction might be best understood by focusing on different processes rather than different systems. Roediger and his colleagues suggested that one way to understand the dissociations that emerged between measures in this literature was to focus on the type of information processing that the measures involved. This work clearly demonstrated that some (but not all) of the variance in the measures was a function not of their "implicitness," but of the nature of the task demands required by the measures.

Similar to this perspective, we have proposed that one important issue is whether a measure taps into biased information processing. In our research we have found that people who show a linguistic bias with regard to African Americans also evaluate a Black but not a White male as more threatening than people who do not show the linguistic bias (von Hippel et al., 1997). In that same article we found that people who show the linguistic bias are also likely to show an attributional bias, which we now refer to as the stereotypic explanatory bias (Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003; Sekaquaptewa & Espinoza, 2004). Sekaquaptewa and her colleagues, in the aforementioned articles, have demonstrated that the stereotypic explanatory bias predicts whether people choose to ask stereotypic questions of Blacks but not Whites in a mock job interview and whether they have negative interactions with Blacks but not Whites in an unstructured setting. It is worth noting, in light of Arkes and Tetlock's (this issue) criticisms, that the negative interactions were a blend of liking by the confederate and nonverbal behavior of the White participant (as reported by the confederate). This combined score had high reliability, but because it was based only on confederate ratings, it does not really address the issue raised by Arkes and Tetlock regarding the ambiguity of nonverbal behavior as an indicator of animosity.

It is our belief that these measures, and similarly derived ones we have developed in the area of attitudes (Vargas, von Hippel, & Petty, 2004) and the self (von Hippel, Larkin, & Shakarchi, 2004) have the potential to supplement the IAT and affective priming procedures to provide a more complete picture of people's unintended cognitive and behavioral responses to oth-

ers. Undoubtedly our measures are also rife with interpretive ambiguities, but because they rely on a very different set of procedures and assumptions, they could prove to be a useful addition to the current implicit measurement menagerie. Additionally, because our procedures are much more deliberative than the IAT and affective priming, they also have the potential to broaden the scope of investigation from automatic stereotyping and prejudice to more thoughtful if not fully conscious processes.

### Conclusions

Arkes and Tetlock (this issue) clearly describe the pitfalls and consequences of the label *implicit prejudice*. They then contrast survey research, which shows great declines in prejudice, with implicit prejudice research, which shows a preponderance of prejudice. Readers are asked to choose which indicator is more valid. This concluding question brings me to my final point as well: Just because we are not sure what the IAT measures does not mean we should accept people's survey responses at face value. First, despite nearly universal self-reported egalitarianism, behavioral measures gathered by testing institutes (such as those sponsored by the Department of Housing and Urban Development) continue to report discrimination in critical areas such as housing, hiring, and bank loans. This tension between self-report and behavior can be seen in the ABC News program "True Colors" when a landlord denies prejudicial motives moments after refusing to show an apartment to an African-American Yuppie that he had just shown to a similar White applicant. As social psychologists, we have a long tradition of believing behavior when it contradicts self-report. Second, a number of studies suggest that among people who typically show no sign of prejudice in self-report or behavior, prejudice can easily rise to the surface when they feel threatened or insecure (Fein & Spencer, 1997). The fact that stereotyping and prejudice automatically manifest themselves among otherwise unprejudiced people in such circumstances (Spencer, Fein, Wolfe, Fong, & Dunn, 1998) suggests that implicit prejudices may indeed be lurking in the hearts and minds of many if not all of us and may indeed be appropriately labeled *prejudice*.

### Note

William von Hippel, School of Psychology, University of New South Wales, Sydney, 2052, Australia. E-mail: w.vonhippel@unsw.edu.au

### References

- Allport, G. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Blair, I. V., Ma, J. E., Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, 81, 828–841.
- Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. Ashmore, L. Jussim, & D. Wilder (Eds.), *Social identity, intergroup conflict, and conflict reduction* (pp. 17–41). New York: Oxford University Press.
- Conrey, R. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2004). *Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance*. Unpublished manuscript, Northwestern University, Chicago.
- Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology*. (pp. 504–553). New York: McGraw-Hill.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.
- Fein, S., & Spencer, S. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73, 31–45.
- Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (in press). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin*.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, 81, 774–788.
- Karpinski, A., & Steinman, R. B. (2004). *Associate strength measures of attitudes: A comparison of the Single Category Association Test to the Implicit Association Test*. Unpublished manuscript, Temple University, Philadelphia.
- Kinoshita, S., & Peek-O'Leary, M. (2004). *Implicit Association Test (IAT): Support for the figure-ground asymmetry account*. Unpublished manuscript, Macquarie University, New South Wales, Australia.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81, 842–855.
- Roediger, H. L. III (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45, 1043–1056.
- Sekaquaptewa, D., & Espinoza, P. (2004). Biased processing of stereotype-incongruity is greater for low than high status groups. *Journal of Experimental Social Psychology*, 40, 128–135.
- Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, 39, 75–82.
- Spencer, S. J., Fein, S., Wolfe, C., Fong, C., & Dunn, M. (1998). Stereotype activation under cognitive load: The moderating role of self-image threat. *Personality and Social Psychology Bulletin*, 24, 1139–1152.
- Vargas, P., von Hippel, W., & Petty, R. E. (2004). Using "partially structured" attitude measures to enhance the attitude-behavior relationship. *Personality and Social Psychology Bulletin*, 30, 197–211.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The Linguistic Intergroup Bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, 33, 490–509.
- von Hippel, W., Lakin, J. L., & Shakarchi, R. J. (2004). *Individual differences in motivated social cognition: The case of self-serving information processing*. Unpublished manuscript, University of New South Wales, Australia.



## Ordinary Forms of Prejudice

**Bernd Wittenbrink**

*Center for Decision Research  
Graduate School of Business  
University of Chicago*

"Why do human beings slip so easily into ethnic prejudice? They do so because the two essential ingredients...—erroneous generalization and hostility—are natural and common capacities of the human mind." (Allport, 1954, p. 17)

One of Allport's (1954) lasting contributions to our understanding of human nature is the recognition that prejudiced attitudes are not necessarily the result of a hateful ideology, or that of a limited intellect, or a disordered personality. Prejudice, instead, may reflect ordinary principles of social psychology: It is the byproduct of basic psychological processes by which the average person understands and relates to the social environment. This interpretation has guided much of the social psychological work on prejudice and intergroup relations ever since Allport's classic text first appeared (e.g., Brewer, 2001; Gaertner, Mann, Murrell, & Dovidio, 1989; Hamilton & Rose, 1980; Park, Judd, & Ryan, 1991; Pettigrew, 1979; Tajfel, 1981). It is also the underlying premise for recent work showing that group attitudes affect people's social perceptions and behaviors implicitly, without a person being aware of such influences, or having control over them (e.g., Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995; Perdue & Gurtman, 1990; Wittenbrink, Judd, & Park, 1997). According to this work, group attitudes and stereotypes stored in long-term memory operate just like ordinary memory contents. To the extent that they are over-learned and frequently accessed in response to a particular stimulus, they are activated automatically whenever the stimulus is present. Several techniques have been proposed that assess a person's propensity to activate group attitudes automatically—among them priming measures (Fazio et al., 1995; Wittenbrink et al., 1997) and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). These measures have received considerable attention from social psychologists as well as from researchers in related fields like political science. In their lead article, Arkes and Tetlock (this issue) resolutely reject these measures. Their rejection is based on a number of reasons that are "part psychological, part philosophical, and certainly part political." Although I do not feel qualified to speak to the philosophical and political aspects of Arkes and Tetlock's criticism, I would like to comment on some of the psychological issues involved. I will focus primarily on what I believe to be the most critical argument raised by Arkes and Tetlock, their assertion that implicit mea-

asures actually do not capture people's attitudes at all. As this argument is not just limited to the specific issue of prejudice but is germane to attitudes in general, my comments address the use of implicit measures for the assessment of attitudes in general.

### The Case Against Implicit Attitude Measures

The argument that implicit measures do not assess attitudes is based on a distinction between two types of associations that a person might have in response to an attitude object: (a) things that a person truly feels and believes and (b) associations that, on further scrutiny, one rejects as being invalid or inappropriate, but that exist merely because of pervasive environmental influences. For example, most people in the United States are familiar with the negative cultural stereotype for African Americans and, thus, have associations stored in memory that link African Americans to stereotypic attributes like *dangerous* and *hostile*. As Arkes and Tetlock (this issue) point out, knowledge of the stereotype, however, doesn't necessarily imply its acceptance. Thus Arkes and Tetlock argue that for associations to be diagnostic of a person's attitude, they have to be *accepted* associations—they have to be "functionally intertwined" with a person's other beliefs and feelings in regard to the attitude object. In contrast, merely *known* associations are extraneous to a person's attitude (for a similar position, see Karpinski & Hilton, 2001).

Implicit measures, by design, assess the associative strength between an attitude object and other evaluatively laden constructs. And although one can frame the assessment context in ways that will make accepted associations more relevant to the response task (see Olson & Fazio, 2004), the measures are in principle sensitive to both accepted and merely known associations. In fact, given that automatic activation is thought to develop from frequent, repetitive experiences with a stimulus (Shiffrin & Schneider, 1977) and given the ubiquitous perpetuation of negative stereotypes about African Americans in the media (e.g., Weigel, Loomis, & Soja, 1980), it is quite possible that culturally shared associations, as opposed to personally accepted associations, play a prominent role in spontaneous evaluations. In contrast, they may be much less relevant for people's attitudes once given an opportunity to reflect on them—which is why Arkes and Tetlock (this issue) conclude that implicit measures are ineffective in assessing prejudice.

### From Associations to Attitudes

Why would we consider implicit measures as evidence for people's attitudes, if they are influenced by factors that people will reject, when given a chance to do so? Before I address this question, let me point out that such discrepancies between different measures of an attitude are not rare and are certainly not limited to comparisons between implicit and explicit attitudes. In fact, the attitude literature is replete with examples in which people's evaluations are influenced by factors that, under different circumstances, they will reject. For instance, one of the classic illustrations of this issue comes from the literature on sexual behavior and related attitudes toward contraceptive use. When people are asked about their attitudes toward condom use, their responses tend to be influenced by cognitive factors, like beliefs about health risks, rather than by affective components, like their desire for instant gratification or feelings of embarrassment. Nevertheless, people's actual behaviors are shaped to a greater extent by affective influences than they anticipate (e.g., Kothandapani, 1971; Marsh, Johnson, & Scott-Sheldon, 2001).

If we applied Arkes and Tetlock's (this issue) argument to this case, we should conclude that affective sources of input are undiagnostic of people's attitudes toward condom use. Because, when given an opportunity, people will reject these influences as inappropriate and as not reflecting their true concerns regarding the potential health risks associated with unprotected intercourse. Of course, such a conclusion is absurd because in the present example these affective sources of input are, if anything, *more* relevant for the evaluations that actually determine people's behaviors than are any of the sources of input that people claim to find acceptable and relevant. Likewise, although people may reject stereotypic associations as invalid and inconsistent with their other beliefs, those rejected associations may nevertheless impact people's evaluations of the group or of individual group members.

The broader point illustrated by this example is that attitudes are based on multiple, and not always evaluatively homogeneous, influences. That is, attitudes are commonly defined as an evaluative tendency, or predisposition, to respond to an attitude object with some degree of favor or disfavor (e.g., Eagly & Chaiken, 1993). In rare cases, such evaluative tendencies are grounded in a single source, like, for instance, a single evaluative association with the attitude object. More commonly, attitudes have multiple sources of input that may not always be consistent in terms of their evaluative implications. With regard to group attitudes, for example, a person is likely to hold many stored associations, of which cultural stereotypes, known members of the group, or personal experiences with members of the group may be some.

The position advocated by Arkes and Tetlock (this issue) ultimately contends that, of all the sources of in-

put that exist, only those will impact a person's evaluative response that are deemed valid and relevant to the attitude object. Therefore, only those explicitly endorsed sources qualify as proper indicators of people's attitudes.

The first problem with this argument is that acceptance of any given influence varies significantly across situations, as the prior example shows and an extensive literature on context effects has shown (for a review, see Sudman, Bradburn, & Schwarz, 1996). Acceptance per se is therefore a poor and ambiguous criterion for determining what kinds of influences a proper attitude measure should actually assess.

The second problem is that in many situations evaluations take place without any consideration about whether they are based on valid and relevant input. In fact, some 20 years of research into the processes that underlie attitudinal responses have firmly established that an evaluation can occur spontaneously, without intent, and without control over or even awareness of its occurrence. Following early demonstrations (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Kunst-Wilson & Zajonc, 1980), many studies now report such spontaneous evaluations, often thought to result from the automatic activation of associated memory contents (e.g., Bargh, Chaiken, Gvender, & Pratto, 1992; Giner-Sorolla, Garcia, & Bargh, 1999; Greenwald, Klinger, & Liu, 1989; Wittenbrink, Judd, & Park, 2001). Spontaneous evaluations occur fast, within a few hundred milliseconds after encountering the attitude object (Fazio et al., 1986). And, as already stated, these early evaluations do not emanate from an intentional, active search for relevant inputs. Instead, they are the result of a passive process that runs its course automatically following exposure to the attitude object (Shiffrin & Schneider, 1977). Because of the passive nature of this process, a person does not even have to be aware of the attitude object or of the evaluation (e.g., Devine, 1989; Greenwald et al., 1989; Wittenbrink et al., 1997). More important, because it is a passive process, the person does not have control over the evaluation, its input or its outcome. In other words, considerations about whether a particular input is actually valid or whether it is relevant to the attitude object are of limited consequence for spontaneous evaluations.<sup>1</sup> Instead, such considerations are part of more deliberate forms of evaluation that can take place following the initial spontaneous evaluation. Whether such deliberate considerations actually take place depends on a variety of factors, like a person's

<sup>1</sup>One way by which such considerations could ultimately impact spontaneous evaluations is if accepted associations become more accessible than other, merely known associations. This may be the case for individuals high in motivation to control prejudice, who may learn to inhibit stereotype activation in response to situational cues that in the past have been associated with prejudiced responses and aversive consequences of those responses (Monteith, Ashburn-Nardo, Voils, & Czopp, 2002).

motivation to spend time and effort on this process, as well as opportunities to actually do so (see Fazio, 1990).

Implicit attitude measures are intended to capture those inputs for spontaneous evaluations that a person will activate habitually in response to a particular attitude object. To the extent that the measures are successful, they should be diagnostic of evaluations in which no further deliberation takes place or in which this deliberation does not affect the final evaluation. Thus, implicit measures, like all other attitude measures, assess evaluative tendencies under specific processing constraints that determine what sources of input can influence the evaluation in what ways. They should be predictive of exactly those evaluations that occur under equivalent processing conditions. As such, implicit measures are not inherently superior measures of attitudes, as it has sometimes been argued. Certainly, there are many contexts in which evaluations are deliberate, and in which, in the case of group attitudes, individuals who reject cultural stereotypes, will try to correct their judgments accordingly (e.g., Wegener & Petty, 1997).

On the other hand, many everyday behaviors are based on simple-minded and superficial evaluations in which the motivation to deliberate is quite limited—such as, for example, when we decide which person to sit next to on a subway train. Likewise, many situations in everyday life place significant cognitive demands on people, as when multiple tasks occur simultaneously or when judgments must be made under time pressure. As a result, a person's capacity for deliberation may often be limited or, in extreme cases, entirely lacking (Bargh, 1997; Correll, Park, Judd, & Wittenbrink, 2002; Gilbert, 1989). In these cases, the input from the initial spontaneous evaluation should be the primary determinant of a person's evaluative response, even though the person may be motivated to reflect on the evaluation in a more deliberate fashion. Implicit measures of attitudes should predict these types of evaluations.

This brief review of the conditions under which spontaneous evaluations occur and under which they may shape people's responses to an attitude object makes clear that Arkes and Tetlock's (this issue) basic distinction between accepted associations and merely known associations is important. In fact, the distinction plays a role in pretty much all recent accounts of how attitudes influence behavior (e.g., Chaiken, 1987; Fazio, 1990; Petty & Cacioppo, 1986; Strack & Martin, 1987; Wegener & Petty, 1997; Wilson, Lindsey, & Schooler, 2000), as well as more general models of behavior and judgment (Sloman, 1996; Smith & DeCoster, 2000; Strack & Deutsch, 2004). However, by no means does this imply that a definition of attitudes—and by implication a definition of prejudice—ought be based on this distinction. Considering only accepted sources of input for an evaluation as in-

dicators of the attitude introduces a fairly arbitrary criterion that varies significantly with context, and more important, it precludes consideration of important aspects of the phenomenon.

With regard to group attitudes in particular, such a limited definition would preclude from consideration an important psychological mechanism that contributes to what Allport (1954) termed the "normality of prejudice" (p. 17). Indeed, for spontaneous evaluations to lead to prejudice it does not require explicit dislike of a particular group of people. All it takes is the acquisition of knowledge as it is perpetuated in the social environment. Although it is important to distinguish such forms of prejudice from more deliberate prejudice, spontaneous evaluations may still lead to quite harmful consequences. It is true that the kind of evaluative tendencies targeted by implicit prejudice measures are not necessarily the kind that will, for example, lead a person to reject an African-American job applicant against all facts of reason. Instead, implicit measures intend to capture evaluative predispositions that will lead to more ordinary forms of prejudice, manifested as, perhaps, the interviewer's silence and lack of encouragement during an interview, which then may ultimately lead to a negative evaluation of the applicant's interview performance.

### Unresolved Issues

Aside from their more principled challenge about whether implicit measures really measure prejudice, Arkes and Tetlock (this issue) also raise several methodological concerns about the specific measurement procedures that have been used. These concerns are justified, although I am optimistic that future research on implicit prejudice measures will eventually resolve them. To date, however, satisfactory answers to several critical methodological questions are still missing. For example, in most cases, the precise mechanism by which implicit measures operate is still not well understood. This is even the case for those measures that have received the most detailed attention—Fazio's (2001) evaluative priming and the IAT (Greenwald et al., 1998). Evaluative priming effects have been explained in terms of spreading activation (Fazio, 2001) as well as in terms of response competition (Klauer & Musch, 2003). For the IAT, an even larger number of explanations exists (e.g., see Brendl, Markman, & Messner, 2001; De Houwer, 2001; Mierke & Klauer, 2001; Rothermund & Wentura, 2004), and the IAT's authors readily acknowledge that research to date "has not yet progressed enough to establish any theoretical interpretation of the IAT effect" (Greenwald & Nosek, 2001, p. 90). Such uncertainty about how implicit measures actually work is obviously problematic. After all, different mechanisms could have different implica-

tions for what precisely it is that these measures assess. At the same time, the empirical evidence on the validity of implicit prejudice measures also remains quite limited. To date, most studies available on this issue are based on small samples, often drawn from college populations, and carried out in laboratory settings with obvious limitations on what sorts of intergroup behaviors can be studied.

Clearly, these are important issues that will have to be addressed to make implicit prejudice measures more than a temporary fad in attitude research. At the same time, however, implicit measures are still quite young. In the relatively short time since their introduction, a substantial body of evidence has emerged. In this growing literature, the methodological issues raised by Arkes and Tetlock (this issue) have not been ignored by researchers in the field but have been addressed head on in three special issues in the field's premier journals, one edited volume in print (Musch & Klauer, 2003), and another one to come (Wittenbrink & Schwarz, in press). The IAT at age six and evaluative priming with just a few years more on its clock do not seem to fare any worse in this regard than most other measurement techniques in social psychology.

### Notes

Many thanks go to Reid Hastie, Chick Judd, and Bernadette Park for their helpful comments on an earlier version of this article.

Bernd Wittenbrink, Center for Decision Research, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637. E-mail: bernd.wittenbrink@gsb.uchicago.edu

### References

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. W., Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition: Vol. 10. Advances in social cognition* (pp. 1–61). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bargh, J. A., Chaiken, S., Gvender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, 62, 893–912.
- Brendl, C., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology*, 81, 760–773.
- Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. D. Ashmore and L. Jussim (Eds.), *Social identity, intergroup conflict, and conflict reduction. Rutgers series on self and social identity* (Vol. 3, pp. 17–41). London: Oxford University Press.
- Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna & J. M. Olson (Eds.), *Social influence: The Ontario Symposium* (Vol. 5, pp. 3–39). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314–1329.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 37, 443–451.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego: Academic.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15, 115–141.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.
- Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology*, 57, 239–249.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman and J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford.
- Giner-Sorolla, R., Garcia, M. T., & Bargh, J. A. (1999). The automatic evaluation of pictures. *Social Cognition*, 17, 76–96.
- Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition*, 17, 35–47.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Für Experimentelle Psychologie*, 48, 85–93.
- Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, 39, 832–845.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, 81, 774–788.
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–50). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, 19, 321–333.
- Kunst-Wilson, W. R., & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, 207, 557–558.
- Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift Für Experimentelle Psychologie*, 48, 161–175.
- Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control pro-



- cesses. *Zeitschrift Fuer Experimentelle Psychologie*, 48, 107–122.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, 83, 1029–1050.
- Musch, J., & Klauer, K. C. (Eds.). (2003). *The psychology of evaluation: Affective processes in cognition and emotion*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, 86, 653–667.
- Park, B., Judd, C. M., & Ryan, C. S. (1991). Social categorization and the representation of variability information. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 211–245). Chichester, UK: Wiley.
- Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, 26, 199–216.
- Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5, 461–476.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration-likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). Orlando, FL: Academic.
- Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology-General*, 133(2), 139–165.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–90.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–31.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8(3), 220–247.
- Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer Verlag.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.
- Tajfel, H. (1981). *Human groups and social categories*. Cambridge, UK: Cambridge University Press.
- Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic.
- Weigel, R. H., Loomis, J. W., & Soja, M. J. (1980). Race relations on prime time television. *Journal of Personality and Social Psychology*, 39, 884–93.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–26.
- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262–74.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology*, 37, 244–52.
- Wittenbrink, B., & Schwarz, N. (Eds.). (in press). *Implicit measures of attitudes: Procedures and controversies*. New York, NY: Guilford.