Annu. Rev. Psychol. 2003. 54:297–327
doi: 10.1146/annurev.psych.54.101601.145225

First published online as a Review in Advance on August 6, 2002

IMPLICIT MEASURES IN SOCIAL COGNITION RESEARCH: Their Meaning and Use

Russell H. Fazio and Michael A. Olson
Department of Psychology, Ohio State University, Columbus, Ohio, 43210-1222; e-mail: Fazio@psy.ohio-state.edu

Key Words attitude measurement, automatic processing, priming, Implicit Association Test

■ **Abstract** Behavioral scientists have long sought measures of important psychological constructs that avoid response biases and other problems associated with direct reports. Recently, a large number of such indirect, or "implicit," measures have emerged. We review research that has utilized these measures across several domains, including attitudes, self-esteem, and stereotypes, and discuss their predictive validity, their interrelations, and the mechanisms presumably underlying their operation. Special attention is devoted to various priming measures and the Implicit Association Test, largely due to their prevalence in the literature. We also attempt to clarify several unresolved theoretical and empirical issues concerning implicit measures, including the nature of the underlying constructs they purport to measure, the conditions under which they are most likely to relate to explicit measures, the kinds of behavior each measure is likely to predict, their sensitivity to context, and the construct's potential for change.

CONTENTS

INTRODUCTION

Over the past few years, there has been a surge of interest in the use of implicit measurement techniques in social psychological research. If this assertion needed any verification, one merely has to call attention to the special issues of four journals recently devoted to the topic: *Journal of Experimental Social Psychology* on "Unconscious Processes in Stereotyping and Prejudice" (Banaji 1997), *Cognition and Emotion* on "Automatic Affective Processing" (De Houwer & Hermans 2001), *Zeitschrift für Experimentelle Psychologie* on "Attitude Measurement Using the Implicit Association Test (IAT)" (Plessner & Banse 2001), and *Journal of Personality and Social Psychology* on "Implicit Prejudice and Stereotyping: How Automatic Are They?" (Devine 2001). The research has involved a variety of domains, including attitudes (e.g., Fazio et al. 1995, Greenwald et al. 1998), stereotypes (e.g., Wittenbrink et al. 1997, Nosek et al. 2002a), self-esteem (Hetts et al. 1999, Bosson et al. 2000, Koole et al. 2001, Rudman et al. 2001b), close relationships (e.g., Banse 1999), and health behavior (e.g., Stacy et al. 1997).

A variety of different implicit measurement techniques have been employed. One such technique involves various priming procedures that have proven useful in the past as a means of assessing what is activated from memory by the presentation of some attitude object (e.g., Gaertner & McLaughlin 1983, Fazio et al. 1986, Greenwald et al. 1989, Perdue et al. 1990). For example, Fazio et al. (1995) examined the consequences of priming participants with photos of black versus white undergraduates. The participants' primary task was to indicate the connotation of an evaluative adjective (e.g., "pleasant" or "awful") as quickly as possible. In the context of a cover story concerning the judgment of word meaning being an automatic skill that should not be disrupted by the participants having to perform an additional task simultaneously, each target adjective was preceded by the brief presentation of a photo. Participants were instructed to attend to these faces so that they would be able to pick them out in a later phase of the experiment. The black and white faces had different consequences for the latency with which participants could indicate the connotation of the subsequently presented target adjective. Relative to what was observed for white faces, black faces facilitated responding to negative adjectives and interfered with responding to positive adjectives. The pattern suggests that, on average, negativity was automatically activated by the black primes.

Probably the most well-known implicit measurement technique is the Implicit Association Test (IAT), developed by Greenwald et al. (1998). This procedure

assesses the strength of an association between a target concept and an attribute dimension by considering the latency with which participants can employ two response keys when each has been assigned a dual meaning. The participants' task is to categorize stimuli as they appear on the screen. For example, in the Greenwald et al. (1998) IAT concerning racial attitudes, participants were first asked to categorize names (e.g., "Latonya" or "Betsy") as typical of blacks versus whites. Here, race is the target concept and the keys are labeled "black" and "white." Participants then categorized a variety of clearly valenced words (e.g., "poison" or "gift") as pleasant or unpleasant, which constitutes the attribute dimension. In the critical phase of the experiment these two categorization tasks were combined. Participants performed this combined task twice—once with one response key signifying black/pleasant and the other labeled white/unpleasant, and once with one key meaning black/unpleasant and the other white/pleasant—in counterbalanced order. The question concerns which response mapping participants find easier to use. In the Greenwald et al. (1998) experiment participants were overwhelmingly faster at responding when black was paired with unpleasant than when black was paired with pleasant. On average, then, the participants found it much easier to associate the target concept black with the attribute unpleasant than with the attribute pleasant.

Many additional implicit measurement techniques have been employed: (*a*) Building upon past research that has examined activation by considering how word fragments are completed (e.g., Warrington & Weiskrantz 1968, Tulving et al. 1982, Bassili & Smith 1986, Gilbert & Hixon 1991), a number of researchers have employed some variant of a word-fragment completion task as an implicit measure of self-esteem (e.g., Hetts et al. 1999), racial prejudice (e.g., Dovidio et al. 1997, Son Hing et al. 2002), or stereotypes (Hense et al. 1995, Sinclair & Kunda 1999).

(*b*) On the basis of work that has demonstrated a linguistic intergroup bias (Semin & Fiedler 1988; Maass et al. 1989, 1996), i.e., a tendency to use more abstract language to describe expectancy-consistent behaviors than expectancy-inconsistent ones, the extent to which descriptions of African American targets' positive and negative behaviors include abstract versus concrete language has been employed as an implicit measure of racial prejudice (e.g., von Hippel et al. 1995, 1997).

(*c*) Based on work indicating that social perceivers tend to explain expectancy-inconsistent events more often than expectancy-consistent events (Hastie 1984), von Hippel and his colleagues (von Hippel et al. 1997, Sekaquaptewa et al. 2002) also have developed an implicit measure of racial stereotyping that focuses on whether a participant's completion of a sentence explains the behavior described in the sentence, as opposed to simply continuing the sentence without explaining. Relatively more such explanations in the case of stereotype-incongruent events serve as an indication of stereotype use on this stereotype-explanatory bias measure.

(*d*) Following from the documented tendency of people to prefer the letters that appear in their own names, especially their initials (Nuttin 1985), the name-letter

preference effect has been employed as an implicit measure of self-esteem (Koole et al. 2001, Jones et al. 2002, Pelham et al. 2002).

(*e*) De Houwer and his colleagues have introduced an affective variant of the spatial Simon task as an implicit measure of attitudes (De Houwer & Eelen 1998, De Houwer et al. 2001a). Participants are required to make some discrimination of the stimuli (e.g., noun/adjective, man-made/natural) but do so by responding "positive" for one category and "negative" for the other. Given that the stimuli themselves vary in their associated valence, this produces both evaluatively congruent trials, in which the valence of the stimulus and its relevant category signal the same response (e.g., saying "positive" to "flower" because it is a noun), and evaluatively incongruent trials, for which response competition exists (e.g., saying "negative" to "happy" because it is an adjective).

(*f*) Finally, Nosek & Banaji (2001) have introduced the Go/No-Go Association Task (GNAT), a variant of the IAT that does not require the use of a contrast category. In the GNAT participants respond to stimuli that represent the target category and the attribute category "good" but do nothing in response to other stimuli. Response latencies or errors are compared to a block of trials for which the participant must respond to items that represent the target category and "bad."

Various physiological approaches have also been employed as implicit measures of attitudes. Vanman et al. (1997) used facial electromyography (EMG) to examine racial prejudice. Both Phelps et al. (2000) and Hart et al. (2000) examined amygdala activation using functional magnetic resonance imaging (fMRI) procedures as an indicant of racial evaluation. Eyeblink startle response to black versus white faces has also been utilized in this way (Phelps et al. 2000, Amodio et al. 2002). Cardiovascular reactivity measures indicative of challenge versus threat have been employed to examine responses to interaction with blacks and other stigmatized individuals (Blascovich et al. 2001). In addition, Cacioppo and his colleagues have employed event-related brain potentials as an on-line measure of the categorization of stimuli as positive or negative (e.g., Cacioppo et al. 1993, Crites et al. 1995, Ito & Cacioppo 2000).

What these various approaches have in common is that they all seek to provide an estimate of the construct of interest without having to directly ask the participant for a verbal report. Their major appeal is that these indirect estimates are likely to be free of social desirability concerns. Often, though not necessarily true for all of the measures, the participant is unaware that attitudes, stereotypes, etc. are even being assessed. The approaches vary in terms of their attempting to assess what is automatically activated in response to a person or object versus employing a presumed consequence of such activation (e.g., influence upon further information processing, as in the stereotype-explanatory bias) as an indirect estimate.

Most of the research that has been conducted has concerned either various forms of priming or the IAT. Hence, our review focuses on these two implicit measures. The burgeoning nature of this literature makes it very difficult to provide a comprehensive review of all the relevant empirical work. As we compiled relevant

articles, we were astounded at how every week was marked by the appearance of new papers. We have made every effort to include recently published and in-press articles. However, to make our task more manageable, we have opted not to include, with very few exceptions, the many relevant conference presentations and unpublished manuscripts that have come to our attention.

Obviously, work on implicit measures relevant to social cognition is in its infancy. The literature, although booming, has not matured to the point at which many firm conclusions can be drawn. When possible, we offer general inferences. More often, however, we offer observations about questions that have arisen as the research has progressed. We hope our structuring the literature into a series of observations or questions that have received, and/or continue to require, attention will prove useful in informing future theory and research.

SOME OPENING OBSERVATIONS

Where's the Theory?

Despite incredible activity, research concerning implicit measures has been surprisingly atheoretical. It largely has been a methodological, empirically driven enterprise. In this review we offer a few observations about issues that have arisen as this area of research has developed. However, readers should be aware that these observations stem from a theoretical perspective that undoubtedly influences how we view and structure this youthful literature. Although it was developed years before this surge of interest in implicit attitude measurement, the MODE model (Fazio 1990, Fazio & Towles-Schwen 1999), we argue, provides a very useful perspective for considering these issues. Moreover, it points to important parallels between the issues that have arisen recently regarding implicit measures and extensively researched questions in classic and current literature regarding attitudes and their influence on judgments and behavior.

The MODE model is one of a class of dual process models (see Chaiken & Trope 1999) that have proven useful in such varied research domains as persuasion, social perception, and attribution. The MODE model's focus on the processes by which attitudes influence judgments and behavior makes it especially relevant to the present concerns. In brief, the model proposes that attitudes can exert influence through relatively spontaneous or more deliberative processes. The former involve judgments of, or behavior toward, an object being influenced by one's construal of the object in the immediate situation—perceptions that themselves can be affected by individuals' attitudes having been automatically activated upon encountering the attitude object. In contrast, deliberative processing involves a more effortful, cost-benefit analysis of the utility of a particular behavior. MODE is an acronym for *m*otivation and *o*pportunity as *de*terminants of whether the attitude-to-behavior process is primarily spontaneous or deliberative in nature. Given the effortful reflection required for deliberative processing, some motivating force is necessary to induce individuals to engage in the processing. The time and the resources to

deliberate—what the model refers to as opportunity—also must exist. In addition to delineating two distinct classes of attitude-behavior processes, the MODE model explicitly postulates the possibility of processes that are neither purely spontaneous nor purely deliberative, but instead are "mixed" processes that involve both automatic and controlled components. Any controlled component within a mixed sequence requires, once again, that the individual be both motivated to engage in the necessary cognitive effort and have the opportunity to do so. Such mixed processes are especially relevant to a number of issues that have received attention regarding implicit measures of attitudes.

Where's the Implicit?

We have to express some misgivings about the very terms "implicit" and "explicit" having been imported from cognitive psychology, at least insofar as they are used to refer to implicit versus explicit attitudes, stereotypes, or self-esteem (Greenwald & Banaji 1995). In cognitive psychology, individuals are said to display implicit memory for a prior event when their performance on some task shows evidence of their having been influenced by that prior event, even though they display no explicit memory for the event; i.e., they report no awareness of the event having occurred (see Schacter 1987, Richardson-Klavehn & Bjork 1988, Roediger 1990). For example, after having been briefly presented with a list of words, individuals might show superior performance on a related word fragment completion task, even though they perform only at chance levels on a recognition task in which they indicate the words they had seen earlier.

If this terminology is to have any similar meaning for attitudes (as well as stereotypes, the self, etc.), then it has to imply that implicit attitudes are ones for which individuals lack awareness. Is that the case? Most importantly, how do we know that individuals lack awareness of their attitudes? Nothing about our current implicit measurement procedures, be it a priming method, the IAT, or one of the other techniques mentioned above, guarantees that participants are unaware of their attitudes. Just because an individual shows evidence that a particular evaluation is primed by the presentation of the attitude object does not mean the individual is unaware of the attitude. Likewise, in the IAT just because an individual has more difficulty associating a given attitude object with the category pleasant than with the category unpleasant does not mean the individual is unaware that he or she views the attitude object negatively. Thus, as long as the term "implicit" is to reflect unawareness, we fail to see any justification for labeling these attitudes as implicit. Discordance between scores on an implicit and an explicit measure should not, in and of itself, be taken as evidence that the implicitly measured construct is an unconscious construct. Yet, the implication that the construct is unconscious can exert subtle, and not so subtle, influences on the way in which we approach theorizing about the antecedents and consequences of the construct.

A second troublesome aspect of the implicit-explicit distinction is that it implies preexisting dual attitudes (or whatever the construct of interest might be)

in memory. That is, if the terms refer to the constructs themselves, then both an implicit and an explicit attitude presumably exist in memory (see Wilson et al. 2000). However, no current explicit measurement procedure guarantees that an explicit attitude is independently represented in memory. Instead, the response to the explicit measure may be constructed on the spot (see Schwarz & Bohner 2001).

For these reasons, it is more appropriate to view the *measure* as implicit or explicit, *not* the attitude (or whatever other construct). What makes priming or the IAT implicit is that these techniques provide estimates of individuals' attitudes without our having to directly ask them for such information. In this sense, the current techniques do not differ from earlier proposals regarding projective methods (e.g., Proshansky 1943) and other unobtrusive measurement procedures (e.g., Webb et al. 1966). Participants may be unaware that their attitudes are being assessed, but that does not mean they are unaware that they possess those attitudes. We would encourage researchers not to equate an implicitly measured construct with an unconscious one. Although an implicit-explicit dissociation may occur because the implicit measure reflects associations to which the individual lacks introspective access, such a dissociation also may occur because people are reluctant to admit (on the explicit measure) to the tendency that is revealed by the implicit measure (see Greenwald et al. 2002a for a related discussion). This ambiguity alone is reason to be wary of the connotations that the term "implicit" carries regarding unawareness. Logically, reference to an implicit attitude, stereotype, or self-esteem should require evidence of unawareness and not solely the use of an implicit measurement technique.[1]

THE RELATION BETWEEN IMPLICIT AND EXPLICIT MEASURES

The literature has included much discussion of the relation between implicit and explicit measures. Are they correlated with one another or are they discordant? Blair (2001), Dovidio et al. (2001), and Brauer et al. (2000) have provided recent reviews. Within the domain of prejudice and stereotypes, the correlations tend to be quite low (e.g., Fazio et al. 1995, Greenwald et al. 1998, Rudman & Kilianski 2000, Kawakami & Dovidio 2001, Monteith et al. 2001, Ottaway et al. 2001, Rudman & Glick 2001, Devine et al. 2002, Dovidio et al. 2002), although there are occasional reports of significant correlations (e.g., McConnell & Liebold 2001, with respect to the IAT; Lepore & Brown 1997; Kawakami et al. 1998; Wittenbrink et al. 1997, with respect to a priming measure). However, some of the initial research using priming methods or the IAT clearly illustrates a correspondence with

[1]Our own preference is for the terms "indirect" and "direct" measures (see Dovidio & Fazio 1992), which are less likely to connote unawareness than are "implicit" and "explicit" (see Richardson-Klavehn & Bjork 1988 for a similar argument). However, the latter terms now appear to be firmly entrenched in the literature.

self-reported attitudes toward mundane, socially noncontroversial objects. When priming techniques were first employed to examine the possibility of automatic attitude activation, attitude objects that participants reported liking or disliking were idiosyncratically selected to serve as the positive and negative primes (Fazio et al. 1986). Similarly, the IAT reveals more positive associations to flowers than to insects, just as people's self-reports indicate (Greenwald et al. 1998). In addition, fairly substantial correlations have been obtained between IAT preference scores and self-reported preferences regarding math versus arts and regarding the 2000 presidential candidates Gore versus Bush (Nosek et al. 2002b).

In our view the variability regarding the correspondence between implicit and explicit measures indicates that discussion of whether a relation exists is not very productive. We already know enough to be able to say that the question has no simple answer. That is, the answer is "it depends." The entire issue is reminiscent of what Zanna & Fazio (1982) referred to as the first generation question regarding the attitude-behavior relation: Is there a relation? Just as in that literature, we need to be asking a "when" question: When, under what conditions, and for what kind of people, are implicit and explicit measures related? The patterns noted above regarding socially sensitive versus nonsensitive issues are consistent with the MODE model (see Dovidio & Fazio 1992, Nosek 2002 for discussions of the role of social sensitivity). The more sensitive the domain, the greater the likelihood that motivational factors will be evoked and exert some influence on overt responses to an explicit measure. The MODE model suggests that the magnitude of the relation between an implicit and an explicit measure will depend on the motivation and opportunity to deliberate. If either motivation or opportunity is relatively low at the time that the explicit response is being considered, then explicit measures should correlate with implicit ones. However, when both motivation and opportunity are relatively high, they are less likely to correlate (see Koole et al. 2001 for evidence on the relevance of opportunity factors to the relation between implicitly and explicitly measured self-esteem).

An unfortunate aspect of this discussion of discordance between implicit and explicit measures is that it appears to have engendered some confusion about the attitude concept. The discussion has raised a conceptual question that we facetiously refer to as "Will the real attitude please stand up?" (e.g., Wilson et al. 2000, Schwarz & Bohner 2001). Is the "real" attitude the one represented by the implicit measure or is it the one reflected in the explicit measure?[2] This question introduces two issues. First, what does "real" mean? There is considerable ambiguity

[2]Fazio et al. (1995) have to bear some responsibility for this confusion, largely because they referred to their unobtrusive, priming technique as the "bona fide" pipeline. However, the conceptual ambiguities also stem from the above-noted general lack of theory in the implicit and explicit measures literature. As we indicate, automatically activated attitudes have a very central role in the MODE model, and "bona fide" is simply intended to indicate that any automatic attitude activation occurs farther upstream than the overt response to an explicit measure.

about this, but to the extent that it refers to a measure's predictive validity, it is important to recognize that both implicit and explicit measures can be predictive of judgments and behavior (Fazio et al. 1995, Dovidio et al. 1997). According to the MODE model, which is more predictive depends on motivation and opportunity to deliberate. When motivation and/or opportunity are low, behavior is expected to be largely a function of the automatically activated attitude, and hence, the implicit measure should prove predictive. When motivation and opportunity are high, the explicit measure should be more predictive—presumably because the explicit measure will have been influenced by these same motivational forces.

The second point to note about this question is that any answer is entirely dependent on one's definition of attitude. A response to an explicit measure is often considered an attitude. Such a response meets most definitions of attitude in that it involves favorable or unfavorable evaluation of the object (see Zanna & Rempel 1988, Eagly & Chaiken 1993 for a discussion of such definitions). However, this view fails to recognize that an explicit measure necessarily involves overt expression of one's attitude. Again from the perspective of the MODE model, such expressions are, for want of any better expression, farther "downstream" than automatically activated attitudes. Responding to an explicit measure is itself a verbal behavior that can be affected by motivation and opportunity, as well as whatever is automatically activated. Any motivational factors that might be evoked can influence the overtly expressed judgment, assuming the situation and behavior in question provide the opportunity for deliberative processing.

PREDICTIVE VALIDITY OF IMPLICIT MEASURES

Priming Measures

Numerous reports attesting to the predictive validity of various forms of priming measures have appeared in the literature. Fazio et al. (1995) found estimates of automatically activated racial attitudes to be predictive of a black target's ratings of the quality of her interaction with the respondent; participants for whom negativity had been activated during the priming task behaved in a less friendly manner when they later interacted with the black target. The attitude estimates also correlated with participants' judgments of the extent to which blacks versus whites were deemed responsible for the escalation in tension and riots that ensued following the verdict of not guilty in the trial of the police involved in the beating of Rodney King. Similarly, Dovidio et al. (1997) observed correspondence between attitude estimates based on a priming procedure and nonverbal behaviors exhibited while interacting with a black and a white interviewer. The more negativity toward blacks that the participants revealed during the priming task, the more frequently they blinked their eyes and the less eye contact they maintained while interacting with the black relative to the white interviewer. In a similar investigation, Dovidio et al. (2002) examined white participants' friendliness while engaged in 3-minute conversations with black and white confederates. Their priming measure correlated

with relative friendliness toward the black as indicated by the confederates' own ratings, judges' ratings of videotapes of the interactions, and judges' ratings of the participants' nonverbal behavior (i.e., silent videos on which only the participant, and not the confederate, was visible).[3]

Wilson et al. (2000) reported a correlation between a priming measure and the number of times white participants touched a black confederate's hand while exchanging a pen they needed to share repeatedly. Jackson (1997) found participants with more negative implicitly measured racial attitudes to judge less favorably an essay purportedly authored by a black undergraduate. Using a priming measure of racial attitudes, Fazio & Hilden (2001) predicted participants' emotional reactions to a public service ad that evoked a seemingly prejudiced response. In an investigation of attitudes toward fat people, Bessenoff & Sherman (2000) found that a priming measure involving photos of fat and thin women predicted how far participants later placed their own chair from that of a fat woman. Fazio & Dunton (1997) observed a relation between estimates of automatically activated racial attitudes and extent to which categorization by race, as opposed to gender or occupation, served as the basis for judging the similarity of pairs of stimulus photos. Dijksterhuis et al. (2000) employed a subliminal priming measure to assess the extent to which individuals associated the "elderly" with "forgetfulness" and found the measure to predict the degree to which participants themselves displayed memory impairment after activation of the category elderly. Finally, in the domain of implicit self-esteem measures, Spalding & Hardin (1999) found that the extent to which the subliminal priming of "me" and "myself" facilitated responding to positive versus negative words predicted participants' apparent anxiety while they were engaged in an interview regarding their own emotional health, but not if the interview concerned their best friends' health.

More complex patterns of findings also have been observed. In particular, a number of investigations have found the relation between automatically activated racial attitudes, as estimated by a priming measure, and subsequent race-related judgments to be moderated by motivational factors. This research, which has employed Dunton & Fazio's (1997) individual difference measure of motivation to control prejudiced reactions, has uncovered such moderating effects with respect to (*a*) responses to the Modern Racism Scale—an explicit measure of racial attitudes (Fazio et al. 1995, Dunton & Fazio 1997), (*b*) evaluations of the "typical Black male undergraduate" (Dunton & Fazio 1997), (*c*) first impression ratings of black target persons (Olson & Fazio 2002a), and (*d*) anticipated comfort in interacting with a black in various unscripted social situations (Towles-Schwen

[3]Interestingly, judges' ratings of the audio channel alone did not relate to the implicit measure but did correlate with an explicit measure, suggesting, as the authors note, that the spoken words stemmed from a more deliberative, motivated process and that the verbal channel provided greater opportunity for exerting such control than did the nonverbal channel. Hence, the data pattern is consistent with the MODE model's predictions regarding the importance of motivation and opportunity (see Fazio & Towles-Schwen 1999).

& Fazio 2002). Consistent with the MODE model, implicitly measured attitudes predicted the various judgments among individuals whose motivation to control prejudiced reactions was relatively low. However, this relation was attenuated as motivation increased. In fact, each of the studies revealed evidence of the relation being reversed among individuals with higher motivation scores. That is, among the highly motivated, those with more negative automatically activated attitudes responded more positively than did those characterized by automatically activated positivity. When the judgmental situation is sufficiently race-related to evoke any motivation to control prejudiced reactions that may characterize individuals, they may respond by overcorrecting for their automatically activated negativity. In terms of Wegener & Petty's (1995) Flexible Correction Model, such individuals possess a naive theory about the biasing effects of their negativity—a theory that overestimates its influence—so they overcompensate. In any case, such findings illustrate the importance of considering motivated processes when examining the predictive validity of an implicit measure. That motivated correction processes can reverse the relation between an implicit measure and overt judgments means that, in situations in which such motivation arises, neglecting to assess motivation to control prejudiced reactions can lead to the erroneous conclusion that the implicit measure is unrelated to the judgment of interest. Instead, the moderating influence of motivational factors may obscure the simple correlation.

In addition to these reports of a priming measure predicting a subsequent judgment or behavior, priming also has proven sensitive when employed as a dependent measure. In an experiment concerning implicit attitude formation via classical conditioning, Olson & Fazio (2002b) had participants undergo a visual task in which one novel stimulus was consistently paired with positively valued stimuli and another with negatively valued stimuli. The names of the two novel stimuli were later presented subliminally as primes during the course of a task in which participants needed to identify the connotation of evaluative adjectives. Responding was faster when the target adjective's valence matched the valence that earlier had been associated with the now subliminally presented prime (see Hermans et al. 2002 for related findings).

Implicit Association Test

A considerable amount of research employing the Implicit Association Test (IAT) has pursued a "known-groups" validity approach, demonstrating that the IAT scores of two groups of individuals differ in the expected way. For example, Japanese-Americans and Korean-Americans have been found to display more positivity toward their respective ingroups (Greenwald et al. 1998), as have East and West Germans (Keuhnen et al. 2001) and Jewish and Christian respondents (Rudman et al. 1999; see Jost et al. 2002 for a system justification perspective on such ingroup bias). IAT effects indicative of ingroup preference have even been found when groups have been created experimentally through a minimal group paradigm (Ashburn-Nardo et al. 2001; see Greenwald et al. 2002b for a related

finding and Otten & Wentura 1999 for similar evidence using a priming measure). Similar differences have been obtained with respect to vegetarians and, although less consistently, with cigarette smokers (Swanson et al. 2001). IAT preferences for white over black have been found to be stronger for white than for black respondents, although the (weaker) preference for white does not appear to reverse itself among blacks (Nosek et al. 2002b; but see Livingston 2002). Teachman et al. (2001) found the IAT to discriminate snake and spider phobics, and Banse et al. (2001) found differences between homosexuals and heterosexuals in a sexual orientation IAT. Correspondence has also been observed between IAT scores reflecting the relative favorability of sodas versus juices and self-reported frequency of drinking the two classes of beverages in the past, and between IAT scores concerning high- versus low-calorie foods and self-reports of past eating behavior (Maison et al. 2001). Expected differences have also been revealed in the manner in which men and women associate gender with mathematics (Nosek et al. 2002a). Recently, Greenwald & Nosek (2001) provided a very useful review of such IAT research.

Additional relations of interest involving the IAT have been observed. In an investigation concerning fMRI-assessed activation of the amygdala, Phelps et al. (2000) found the strength of amygdala activation in response to unfamiliar black versus white faces to covary with race preference as assessed by the IAT. Olson & Fazio (2001) found the IAT to be sensitive to the implicit formation of attitudes via classical conditioning. Finally, in a study concerning the strength of association between the target concept "gender" and the attribute categories "career" versus "household," Gawronksi et al. (2002) found IAT scores to relate to memory performance in the "who said what?" paradigm. When unable to retrieve sufficient information from memory, participants with stronger stereotypic associations exhibited a bias toward assigning statements in a stereotype-consistent manner.

In contrast to the numerous investigations concerning known-group differences, less work has been conducted concerning the prediction of behavior from IAT scores. The evidence that does exist is mixed. Disappointingly, Karpinski & Hilton (2001) observed no relation between IAT scores reflecting relative preferences for apples versus candy bars and participants' subsequent choice behavior. However, Greenwald & Farnham (2000) did obtain some evidence suggesting that an IAT-based measure of self-esteem predicted reactions to success versus failure experiences; individuals with higher scores on the implicit self-esteem measure tended to be less affected by the feedback. Similarly, Jordan et al. (2002) reported a positive relation between a self-esteem IAT and a behavioral measure that involved persistence in the face of failure. In addition, Rudman & Glick (2001) found a correlation between an IAT assessment of the association between gender and agency versus communality and judgments of the social skills displayed by a highly agentic female applicant for a job that required social sensitivity. However, no significant correlation was observed when the job description did not necessitate sensitivity to others or with respect to ratings of the hireability of the applicant.

More promising are results from an intriguing and sophisticated investigation conducted by McConnell & Liebold (2001). After a brief interaction with a white

experimenter, participants completed a set of questionnaires that included various explicit measures of prejudice, followed by a racial-prejudice IAT. They then interacted with a black experimenter. Both experimenters provided ratings of the interaction, as did judges who examined videotapes of the interactions. Impressively, IAT scores related significantly to the experimenters' and the judges' molar ratings of the interaction with the black versus the white experimenter. In addition, IAT scores indicative of prejudice against blacks were also associated with a variety of micro-behaviors including less speaking time, less smiling, fewer extemporaneous social comments, more speech errors, and more speech hesitations in the interaction with the black (versus white) experimenter.

Although these findings are unquestionably promising, a note of caution is required. As the authors indicated, the confirming results may have been facilitated by the sequence of events. In particular, the interaction with the black experimenter was immediately preceded by the administration of the explicit measures and the IAT, which would have resulted in the conscious activation of racial attitudes. Heightened attitude salience has been shown to increase attitude-behavior consistency (Snyder & Swann 1976, Snyder & Kendzierski 1982). Completion of the IAT also may have encouraged categorization of the second experimenter as "black" instead of as a member of some other applicable social category, which again may have promoted attitudinally consistent behavior.

Such an interpretive issue should not be viewed as calling for administration of the IAT after the behavioral assessment. This order of events merely poses a different set of issues. Instead of illustrating predictive validity, such a relation may be due to an effect of a recently performed and hence salient behavior on the IAT. Researchers interested in the attitude-to-behavior relation in the 1970s and 1980s were very sensitive to this self-perception possibility. This concern remains applicable even when attitudes are measured implicitly instead of explicitly. Indeed, some impressive findings regarding correlations between behavior and the IAT are rendered ambiguous by this possibility. For example, participants in an intriguing study by Lemm (2001) were interviewed about their attitudes toward gay men. Midway through the interview, the interviewer was identified as gay for half the participants. The interviewer rated the participants' apparent comfort level after each of the two interview segments. The change in rated comfort correlated significantly with an IAT involving the target categories gay versus straight. However, the IAT was administered immediately after the interview and, hence, participants' construals of a gay may have been affected by their recent experience. Similarly, a recent study by Asendorpf et al. (2002) involved an interaction with a physically attractive stranger of the opposite sex. After the conversation, participants completed an IAT assessing the strength of the association between "me" and "shy." These IAT scores correlated with coders' global ratings of the shyness exhibited during the interaction, as well as with more micro-ratings of such nonverbal behaviors as body tension. However, as the authors note, it is conceivable that this relation reflects the participants' recent self-observation, and not the IAT's having tapped an enduring and behaviorally predictive trait.

Clearly, more research on the predictive validity of the IAT is needed. Especially informative will be work that administers the IAT prior to the collection of the dependent measure, so as to avoid any possibility that the behavior is influencing the attitude scores, but that also involves some separation in time between IAT administration and the behavior of interest. Indeed, a more optimistic outlook regarding the predictive validity of the IAT is provided by a study by Rudman & Lee (2002), which involved a racial prejudice IAT administered in a separate, earlier session. In the second session participants were exposed to either violent and misogynous rap music or popular music, which served as a priming manipulation of racial stereotypes. A subsequent person-perception task involved the participants evaluating an ambiguously described black or white target. Although the sample was too small for the correlations to differ significantly across conditions, a significant correlation between IAT scores and evaluations of the target was evident only in the condition which involved a black target and prior exposure to rap music.

Some evidence suggests that the ability of the IAT to predict judgments and behavior may sometimes be moderated by motivational factors—just as is true for priming procedures. Florack et al. (2001) observed a moderating influence of need for cognition on the relation between an IAT reflecting favorability toward Turks, relative to Germans, and subsequent judgments of a Turkish juvenile delinquent. The expected relation was most apparent among individuals less inclined to engage in effortful cognitive processing. Hence, future research concerning the predictive validity of the IAT may benefit from the consideration of moderating variables.

Other Measures

Although less widely examined than either priming measures or the IAT, other implicit measures have proven predictive of relevant judgments and behavior. For example, using the linguistic intergroup bias measure mentioned above, von Hippel et al. (1997) found a relation between the implicit measure and the extent to which participants judged a videotaped African-American requesting money from another person as threatening. Similarly, the tendency to engage in the earlier described stereotype-explanatory bias has been found to predict the quality of subsequent social interaction (Sekaquaptewa et al. 2002). Spontaneously generating relatively more internal attributions (or fewer external attributions) for a black's stereotype-inconsistent behaviors than stereotype-consistent behaviors was predictive of relatively more positive interactions in a condition involving a black confederate than in one involving a white. Using a word fragment completion task to implicitly measure racial prejudice toward Asians, Son Hing et al. (2002) identified as "aversive racists" those individuals with low-prejudiced scores on an explicit measure but high-prejudiced scores on the implicit measure. Relative to participants who scored low on both the implicit and the explicit measures, these aversive racists exhibited more negative feelings and less discriminatory behavior after having been made to feel hypocritical about the extent to which

they lived up to their egalitarian values. Finally, in the self-esteem domain, Hetts et al. (1999) found the positivity of word fragment completions that followed reading an individualistic statement versus a collectivistic statement to relate to the recency of Asian-Americans' immigration to the United States. (Similar findings were obtained in an experiment that involved response latencies to positive and negative target words that had been preceded by the primes "me" versus "us.")

QUESTIONABLE INTERRELATIONS AMONG IMPLICIT MEASURES

One of the most disturbing trends to emerge in the literature on implicit measures is the many reports of disappointingly low correlations among the measures. Bosson et al. (2000) observed null relations among a variety of implicit measures of self-esteem, including the IAT, supraliminal and subliminal priming, and name-letter preference. Brauer et al. (2000) observed a small, albeit significant, correlation of .27 between two versions of a priming measure of attitudes toward women, one based on a lexical-decision task and one based on an adjective-connotation task. Somewhat more promising are the findings obtained by Rudman & Kilianski (2000) in a study that employed as primes schematic drawings of males and females in high and low authority roles. Participants also completed an IAT examining the association between the target concept "gender" and the attribute categories of high- versus low-status occupational roles. IAT scores correlated significantly (ranging from .27 to .38) with a series of contrasts reflecting negativity in response to the high authority female primes relative to various combinations of the other primes. However, no such relations were observed with respect to IATs assessing the association between gender and career versus domestic categories and between gender and agentic versus communal qualities. Additional reports of null relations between IAT and priming measures include the work of Marsh et al. (2001) concerning attitudes toward condom use and Sherman et al. (2002) concerning attitudes toward cigarette smoking. In our own lab we have repeatedly failed to observe correlations between IAT measures and priming measures of racial attitudes ($\underline{r}$'s ranging from $-.13$ to .05 across four studies).

Unquestionably, part of the problem with these disappointing correlations among various implicit measures is their rather low reliability. The problem is certainly not unique to the field of social cognition; it has been noted by cognitive psychologists interested in implicit memory measures as well (e.g., Perruchet & Baveux 1989, Buchner & Wippich 2000). Test-retest reliability for the IAT does tend to reach a respectable level of .6 or higher (Bosson et al. 2000, Greenwald & Nosek 2001), as also is true for the name-letter preference task (Bosson et al. 2000). However, the few reports regarding test-retest reliability for various priming measures have ranged from abysmally low (Bosson et al. 2000) to moderate levels of $\sim$.5 (Kawakami & Dovidio 2001). Such test-retest coefficients are likely

to vary considerably as a function of a number of factors. Too few trials are likely to produce inadequate and unstable estimates; we note, for example, that the supraliminal priming measure employed by Bosson et al. included a mere 20 trials, only 4 of which were involved in the calculation of the implicit score. On the other hand, a large number of trials is likely to promote boredom and fatigue, and this problem will grow only more serious the second time the task is performed.

To date, the most sophisticated examination of measurement error and the interrelations among various implicit measures has been provided by Cunningham et al. (2001), who administered each of three implicit measures of racial prejudice on four separate occasions. The measures were (*a*) a response-window version of a supraliminal priming procedure, in which black and white faces served as primes, and participants were required to indicate the valence of a target word before a very brief deadline (a brief response window greatly increases the number of errors made, and changes the focus of the analyses from latencies to error rates) (see Draine & Greenwald 1998); (*b*) a typical race-related IAT involving black and white faces and positive and negative words; and (*c*) a response window version of this same IAT. Although the average test-retest correlation was only .27 and the average intercorrelation only .19, separating measurement error from estimates of stability via a latent variable analysis substantially improved the interrelations. For example, the priming measure now correlated .55 and .53 with the typical and response window versions, respectively, of the IAT.

Measurement error, then, unquestionably plays a role in the low relationships that have been observed among various implicit measures. However, it is not clear that it comprises the full story. The Cunningham et al. (2001) study involved intensive and repeated administration of multiple implicit measures, as well as an explicit measure—all of which should have enhanced participants' awareness that race was the focus of the research. As a result, and as we argue below, responses to the faces presented during the priming task may have been altered from what is typical in a priming measure. Moreover, we have to admit to doubts about what any sophisticated correction for measurement error can accomplish when the interrelation between a priming measure and the IAT are as low as has been reported in some of the work noted above—essentially zero. To better understand any interrelations that might or might not exist, it is important to consider the mechanisms that underlie the various measures.

UNDERLYING MECHANISMS

Careful consideration of the mechanism underlying any given implicit measurement procedure can provide invaluable information about its focus and its similarity to other measures. De Houwer (2002) provided an intriguing structural analysis of a variety of indirect measures of attitude. Here, we focus on a consideration of only the priming and the IAT techniques.

What Drives Priming?

Largely because decades of research in cognitive psychology and, more recently, social psychology have employed priming paradigms, the mechanisms underlying priming measures are fairly well understood. Such implicit measures focus on what is automatically activated by the primed stimulus. Automatic activation of the evaluation associated with a prime produces a processing advantage for evaluatively congruent targets. Both an encoding advantage and a response-competition mechanism have been implicated (see Fazio 2001 for a review). When the task involves word naming, any facilitation produced by the prime is viewed as the result of activation spreading from the prime to the associated evaluation, thus diminishing the amount of additional activation required for the target to reach the threshold necessary for a response (e.g., Bargh et al. 1996, De Houwer et al. 2001b). When the task concerns judging the evaluative connotation of the target, response competition/facilitation is also involved. The evaluation activated by the prime readies the participant to respond when the subsequently presented target is evaluatively congruent, but the response suggested by the prime must be inhibited in order to respond accurately to an evaluatively incongruent target (e.g., Klauer 1998, Wentura 1999, Klinger et al. 2000). In either case, the process is potentiated by the automatic activation of an evaluation associated with the prime, and any consequent effect on the latency of response to positive versus negative targets provides information about the evaluation of the prime.

When priming procedures are employed as an implicit measure, exemplars of the category of interest typically are presented as primes. The average response across the set of exemplars is treated as a measure of attitude toward the group as a whole. For example, photos of blacks and whites might be employed as a means of estimating racial attitudes. Hence, the validity of any priming measure is likely to depend on the representativeness of the stimuli chosen to serve as primes. Indeed, recent findings by Livingston & Brewer (2002) illustrate that different effects can emerge for photos of African-Americans that vary in prototypicality. Whereas the typically observed effects reflecting greater negativity in response to black faces than white faces were replicated when prototypical black faces were employed, this was not true of a set of photos judged to be less prototypical, even though the faces were definitively identifiable as African-American. Once again pointing to parallels with earlier literature concerning attitudes, it is worth noting that considerable evidence exists regarding the role of prototypicality as a moderator of the relation between attitudes toward a general group and behavior toward a specific group exemplar (e.g., Lord et al. 1984; see Lord & Lepper 1999 for a review).

What Drives the Implicit Association Test?

The assumption that forms the basis for the IAT is straightforward: "... if two concepts are highly associated, the IAT's sorting tasks will be easier when the two associated concepts share the same response than when they require different

responses" (Greenwald & Nosek 2001, p. 85). Precisely how the IAT works remains unclear, however. What is clear is that an early concern regarding the potential confounding of familiarity with black versus white names in racial IATs, as in Greenwald et al. (1998), is not a serious problem. Since then, researchers have observed IAT effects even after carefully matching stimuli and/or statistically controlling for familiarity (Rudman et al. 1999, Dasgupta et al. 2000, Ottaway et al. 2001).

A number of possible mechanisms for the IAT have received attention recently. Included are (*a*) the idea that incompatible response mapping produces a *shift in response criteria*, resulting in slower responding to both target and attribute stimuli on such incompatible mapping trials (Brendl et al. 2001); (*b*) a *figure-ground asymmetry* model that focuses on participants perceiving one response category as figure on the ground of the opposing response category (Rothermund & Wentura 2001); and (*c*) a *task-set switching* account that focuses on the possibility of respondents neglecting the instruction to switch tasks (assigning stimuli via the attribute categories versus the target categories) during the compatible mapping phase of the IAT, because responding on the basis of the attribute-related information is sufficient to meet the goals of fast and accurate responding (Mierke & Klauer 2001).

Especially informative, in our view, is De Houwer's (2001) consideration of what he refers to as "relevant" and "irrelevant" feature accounts of the IAT. De Houwer noted that the structurally relevant feature of any given target concept stimulus (e.g., "tulip") in the IAT is its membership in one of the target categories ("tulip" is a flower, not an insect). However, one can also maintain that IAT performance is sensitive to what is structurally an irrelevant feature—the valence of individual target stimuli ("tulip" itself is positive). De Houwer (2001) pointed out that the typical IAT perfectly confounds the relevant and irrelevant feature possibilities; all exemplars of the positive target category (e.g., flowers) are themselves positive, and all exemplars of the negative target category (e.g., insects) are themselves negative. To examine whether IAT performance is a function of the valence of the target categories (the relevant feature account) or the exemplars (the irrelevant feature account), he conducted an experiment with British participants that involved the target categories "British" and "foreign" but included both positively valued (e.g., Princess Diana) and negatively valued (e.g., the name of a well-known mass murderer) British names, and positively valued (e.g., Albert Einstein) and negatively valued (e.g., Adolf Hitler) foreign names as the exemplars. Performance was found to be affected only by whether "British" and "positive" were represented by the same or different keys. Contrary to the irrelevant-feature account, whether the individual exemplar was itself positively or negatively valued had no effect. Thus, IAT performance is very much dictated by the task instructions to categorize the target stimuli. The task requires such categorization, encouraging construal of the exemplars in terms of the target categories to such an extent that processing of these stimuli is limited to the relevant task feature—their category membership. This inference is further supported by research conducted by Mitchell et al. (1999), who found different IAT scores for the same stimuli

when the names of famous black athletes and white politicians (e.g., Michael Jordan, Jesse Helms) were categorized as black versus white or as athletes versus politicians.[4]

Implications

The importance of these findings is the suggestion that, in contrast to priming measures, the IAT has little to do with what is automatically activated in response to a given stimulus. Although IAT effects are often referred to as "automatic preferences," this use of the term automatic appears to have a very different meaning than it does in the context of priming procedures. In contrast to the latter case, in which the term concerns the spontaneous activation of an evaluation in response to the primed stimulus, for the IAT the emphasis is on controllability: "IAT responses are considered automatic because they are expressed without intention or control, although perceivers may become aware of the attitude under scrutiny during the task" (Dasgupta et al. 2000, p. 317). The distinction is important, and as De Houwer's (2001) findings imply, the IAT seems to assess associations to the category labels, not automatically activated responses to the individual exemplars. Thus, in the context of race, whereas priming procedures provide an estimate of the average evaluation evoked by the black (versus white) faces, the IAT provides an estimate of the strength of association between the category label "black" (versus "white") and negativity.[5] The potential exists for these two estimates to differ, especially in any situation that does not promote categorization of a given person by race.

Its operation at the level of the category also means that the IAT may be influenced by associations other than those involved in a perceiver's own automatically activated response to a given exemplar—ones that are potentially independent of the association between a perceiver's own evaluation and the category in question. Indeed, Karpinski & Hilton (2001) have made exactly this argument. As an explanation for their finding that an IAT revealed an average preference for apples over

[4]This is not to say that IAT performance cannot be affected by the specific stimuli that are presented. Indeed, De Houwer (2001, footnote 4) summarized an experiment involving an atypical IAT—one that used evaluatively neutral ("person" and "animal") instead of evaluatively laden target categories. In this case, valence of the individual stimulus items did matter. Moreover, Steffens & Plewe (2001) found that the nature of the items reflecting the attribute dimension can affect IAT scores. Female participants more easily associated "female" with "pleasant" when the pleasant items were more stereotypic of females than males. However, this finding does not speak to whether the valence of the exemplars representing the target categories affects response latencies. The major point we wish to make is not that IATs in which performance is affected by stimulus items cannot be designed, but that the IAT has little to do with the automatic activation of evaluations in response to the target exemplars.

[5]This distinction would not apply for priming measures that involve the subliminal presentation of the words "black" or "white" as primes (e.g., Wittenbrink et al. 1997).

candy bars (a preference that was not apparent on an explicit measure of liking), these researchers noted: "In our society, there are an abundance of positive associations and virtually no negative associations with apples. For candy bars, however, the messages are much more mixed" (p. 783). They suggested that the IAT may be influenced by such environmental associations. To test this idea, Karpinski & Hilton exposed research participants to a large number of word pairings associating the word "youth" with various positive items and the word "elderly" with negative items or vice versa. When the pairings were contrary to greater favorability toward youth, the extent of the preference participants revealed for youth on a subsequent IAT was reduced. In contrast, explicit measures were unaffected.

Such extrapersonal associations may play an important role in IATs that assess racial prejudice. It is possible that even individuals for whom negativity is not automatically activated in response to black exemplars, as assessed by priming, may possess and perceive strong associations between the category "blacks" and negativity; they may recognize, for example, that blacks have been historically portrayed in a negative manner by American society. This knowledge may easily come to mind when the IAT presents them with a black + negative response mapping and may facilitate their responding. This may be why the IAT, relative to priming measures, typically reveals such a higher percentage of a white college student sample to be negative toward blacks. Extrapersonal associations may also bear some responsibility for the finding that black respondents often display negativity toward blacks on the IAT (Nosek et al. 2002b).

However, both Banaji (2001) and Lowery et al. (2001) have persuasively articulated the difficulty of distinguishing cultural associations from personal ones, i.e., of separating the individual from the culture. Where do attitudes originate if not from learning experiences within one's culture? That self and culture are inextricably intertwined is undeniable. However, individual learning experiences certainly can produce an evaluation that differs from what is modal in the culture. An individual with an allergy to peanuts unquestionably develops a strong negative association, one that is easily capable of automatic activation upon reading of their presence in an entree's menu description, yet the individual also possesses the knowledge that most people like peanuts. In a priming assessment of the individual's attitude, photos of peanuts, peanut butter, etc. would probably produce evidence of automatic activation of negativity. Given that the IAT seems to reflect not what is automatically activated when a stimulus is presented but associations to the category response label (in this case, peanuts), it is quite possible that the two measures will produce diverging estimates of our allergy sufferer's attitude. Especially when faced with a response mapping that forces pairing of peanuts and positive, the individual may "recruit" the knowledge of others' generally positive evaluations of peanuts to facilitate the goal of responding quickly. In other words, extrapersonal knowledge may assist the individual in solving the response mapping problem posed by the IAT. If so, the implicit score is affected by knowledge that would not have been automatically activated upon the presentation of peanuts, and hence, some degree of disparity would be observed between the priming

and IAT estimates. Moreover, to the extent that this happens more for some peanut-allergy sufferers than others, the two measures may yield different rank orderings of the individuals. Any such disparities, however, should be limited to domains (or individuals) characterized by diverging personal and environmental associations.

In sum, then, consideration of the mechanisms underlying priming and IAT procedures appears to inform issues regarding their predictive validity and interrelation. Some possibilities arise that, although admittedly speculative, are clearly worthy of future research. The less representative the sample of category exemplars used as primes in the priming procedure, the less valid the estimate of attitude toward the category and also the lower the correspondence with the IAT. In addition, the more variability that exists in evaluative reactions to category exemplars, the less meaningful the average reaction across exemplars becomes and the more likely it is that the mean response will differ from the category evaluation provided by the IAT. For example, for most people, more variability probably exists regarding their evaluation of specific cars than of specific flowers. Moreover, the category "cars" may evoke thoughts of pollution, overcrowded highways, etc. that are not evoked by the presentation of specific cars. Ultimately, the predictive utility of the two methods may depend on the nature of the judgment or behavior one is attempting to predict. Because its focus concerns evaluative associations to the category, the IAT may be superior for predicting behaviors at the category level (e.g., support for a category-related social policy proposal). A priming measure may be superior when predicting behavior toward an exemplar of the category (e.g., judgments of a category member).[6]

EFFECTS OF CONTEXT ON IMPLICIT MEASURES

Explicitly reported evaluations are known to be context dependent. That is, they depend on one's construal of the object in the immediate situation and on the comparison standards employed to make the judgment (e.g., Bruner 1957, Sherif & Hovland 1961, Carlston & Smith 1996, Higgins 1996). Considerable recent research demonstrates that implicit measures also display some sensitivity to context (see Blair 2002 for a detailed review). For example, in a priming procedure, Wittenbrink et al. (2001a) obtained evidence of less automatically activated negativity in response to black faces when those faces were presented within the background context of a church interior as opposed to an urban street corner. Related effects have been observed as a function of individuals' motivational states.

[6]IAT scores may predict behavior toward an exemplar to the extent that the situation promotes categorization by race. We note that IAT-behavior correlations were found when the sequence of events made race salient (McConnell & Liebold 2001) and when stereotypically black music, but not other music, preceded trait ratings of a black target (Rudman & Lee 2002).

Sherman et al. (2002) found a priming measure to be sensitive to heavy cigarette smokers' experimentally manipulated level of deprivation; the greater the deprivation, the more positivity these smokers experienced in response to cigarette-related photos. Similarly, Pratto & Shih's (2000) priming measure revealed more evidence of prejudice toward an outgroup on the part of participants characterized by higher social dominance orientation after the status of their ingroup had been threatened. Thus, the evaluations that are automatically activated in response to a given stimulus depend on one's construals of the stimulus in a given context. How a multiply categorizable object or target person is construed in any given situation will determine the attitude that is automatically activated (see Fazio 1986, Smith et al. 1996, Fazio & Dunton 1997 for discussion of such categorization processes).

Similar phenomena have been demonstrated multiple times with respect to the IAT. Essentially, pre-IAT exposure to different kinds of information appears to influence participants' construal of the category labels involved in the IAT. Salient positive information about a given category produces IAT scores indicative of more favorable evaluations of the category. Such manipulations have involved exposure to movie clips that depicted blacks at a harmonious family barbecue versus an argumentative, gang-related scene (Wittenbrink et al. 2001a), presentation of a series of either admired black individuals and disliked whites or disliked blacks and admired whites (Dasgupta & Greenwald 2001), assigning participants to interact with a black partner who occupied a superior or subordinate task role (Richeson & Ambady 2002), exposure to violent and misogynous rap music (Rudman & Lee 2002), and introduction to the IAT (specifically as a procedure for assessing prejudice) by a black versus a white experimenter (Lowery et al. 2001). Finally, in research involving implicit measurement of gender stereotypes, Blair et al. (2001) found that having participants engage in counterstereotypical mental imagery (i.e., imagine a strong woman) reduced the strength of stereotypic associations.[7]

ADDITIONAL QUESTIONS FOR FUTURE RESEARCH

Role of Awareness

We argued above that an implicitly measured attitude should not be assumed to be nonconscious. Our own research regarding the moderating role of motivation to control prejudice would lead us to speculate that people generally are aware

[7]These various manipulations of context bear some similarity to the earlier-described manipulation of environmental associations employed by Karpinski & Hilton (2001). We find it puzzling that similar evidence has been interpreted both as problematic for the IAT and as a presumably advantageous sign of the IAT's sensitivity to context. In our view the critical issue to be resolved is whether such contextual manipulations (*a*) produce an actual change in the relevant mental representation, to which the IAT is sensitive, or (*b*) simply render momentarily salient positive or negative construals of a given category label, which then assist the participant with the response-mapping problem posed by the IAT.

of their automatically activated racial attitudes. In a variety of studies, the more motivated show evidence of having "corrected" for their automatically activated attitudes. Given that models of correction processes (e.g., Wegener & Petty 1995) require that individuals be aware of a potential bias in order to engage in effortful correction, the implication is that people are aware of their automatically activated racial attitudes. However, such indications of awareness may not arise in other domains, and we do not question that it is possible for people to possess, and be influenced by, attitudes of which they are unaware. In any case, the issue of awareness needs to be addressed more directly.

Stereotype Versus Attitude Activation

Just what associations are activated in response to a given prime: a stereotype and/or an attitude? Wittenbrink et al. (2001b) suggested that it can be either, depending on the nature of priming task; they obtained some evidence suggesting that a priming measure based on a lexical decision is more sensitive to stereotype activation, whereas one based on adjective connotation is more sensitive to attitude activation. Conceptually, however, this suggests an adaptive system that is responsive to what information may prove most functional in terms of meeting the needs of the situation. Hence, stereotypes may be activated only given a relevant processing goal (e.g., Macrae et al. 1997, Livingston & Brewer 2002). Certain decision situations may require some sense of the extent to which a person or object matches some ideal set of attributes, and such a need for attribute information may evoke stereotypes. Other decisions are driven more purely by evaluation and, hence, may be unlikely to evoke associated attribute information, but very likely to evoke associated attitudes.

Changing Automatically Activated Constructs

Literature on the development of automatism points to the importance of consistent and repeated mapping (Schneider & Fisk 1982, Schneider & Shiffrin 1977). Hence, it is thought that such attitudes develop slowly and that any change will be difficult and will require extensive practice. Kawakami et al. (2000) demonstrated that hundreds of trials of overt rejection of the stereotypic association were necessary to produce change in later stereotype activation. Rudman et al. (2001a) observed changes regarding both prejudice (on an IAT) and stereotyping (a priming lexical decision measure) among students enrolled in a semester-long prejudice and conflict seminar. However, other indications in the literature provide a reason for a more optimistic outlook regarding the potential for change that stems, not from negation or heightened sensitivity, but from the creation of new counterstereotypical or counterattitudinal associations. Recall the research presented earlier regarding the malleability of implicit measures. Exposure to counterexemplars reduced the extent of prejudice apparent on the IAT. Moreover, this reduction continued to be observed when participants returned 24 hours later for re-administration of the IAT (Dasgupta & Greenwald 2001). Whether this reflects the learning of a strategy

by which participants approached the IAT specifically or an actual change in the associations that might be automatically activated remains to be seen. Whether any such change would persist over extended time and generalize across situations is also unknown. Nevertheless, the various malleability findings imply that appropriate environmental pairings have the potential to counter the associations that have been learned in the past.

FINAL THOUGHTS

Although obviously in its infancy, social cognition research on implicit measures is certainly progressing. One must be impressed by the sheer amount of empirical work that has been conducted in such a short time. As our knowledge base accumulates and our understanding of the mechanisms underlying various implicit measurement techniques increases, the questions that we can ask will become sharper and the contributions to basic theory all the stronger. We are convinced that, when their application, use, and interpretation is guided by relevant theory and past literature, implicit measures have the potential to serve as useful methodological tools for testing hypotheses. Although we have no confidence that all of our speculative commentary will remain applicable as this literature burgeons even further, it is our hope that our efforts here to structure the empirical findings accumulated to date will facilitate such growth and development.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by Senior Scientist Award MH01646 and Grant MH38832 from the National Institute of Mental Health. The authors thank the following individuals for their helpful feedback on an earlier version of the manuscript: Hal Arkes, Marilynn Brewer, Tim Brock, Jan De Houwer, Dick Eiser, Susan Fiske, Tony Greenwald, Kristi Lemm, Brian Nosek, Rich Petty, Laurie Rudman, Jim Sherman, Eric Uhlmann, Bill von Hippel, and Mark Zanna.

LITERATURE CITED

Amodio DM, Harmon-Jones E, Devine PG. 2002. Individual differences in the activation and control of affective race bias as assessed by startle eyeblink responses and self-report. *J. Pers. Soc. Psychol.* In press

Asendorpf JB, Banse R, Mücke D. 2002. Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *J. Pers. Soc. Psychol.* 83:380–93

Ashburn-Nardo L, Voils CI, Monteith MJ. 2001. Implicit associations as the seeds of intergroup bias: How easily do they take root? *J. Pers. Soc. Psychol.* 81:789–99

Banaji MR, ed. 1997. Unconscious processes in stereotyping and prejudice. *J. Exp. Soc. Psychol.* 33:449–560

Banaji MR. 2001. Implicit attitudes can be measured. In *The Nature of Remembering:*

Essays in Honor of Robert G. Crowder, ed. HL Roediger, JS Nairne, pp. 117–50. Washington, DC: Am. Psychol. Assoc.

Banse R. 1999. Automatic evaluation of self and significant others: affective priming in close relationships. *J. Soc. Pers. Relat.* 65:272–81

Banse R, Seise J, Zerbes N. 2001. Implicit attitudes toward homosexuality: reliability, validity, and controllability of the IAT. *Z. Exp. Psychol.* 48:1145–60

Bargh JA, Chaiken S, Raymond P, Hymes C. 1996. The automatic evaluation effect: unconditional automatic attitude activation with a pronunciation task. *J. Pers. Soc. Psychol.* 62:893–912

Bassili JN, Smith MC. 1986. On the spontaneity of trait attribution: converging evidence for the role of cognitive strategy. *J. Pers. Soc. Psychol.* 50:239–45

Bessenoff GR, Sherman JW. 2000. Automatic and controlled components of prejudice toward fat people: evaluation versus stereotype activation. *Soc. Cogn.* 18:329–53

Blair I. 2001. Implicit stereotypes and prejudice. In *Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition*, ed. GB Moskowitz, pp. 359–74. Mahwah, NJ: Erlbaum

Blair I. 2002. The malleability of automatic stereotypes and prejudice. *Pers. Soc. Psychol. Rev.* 6:242–61

Blair I, Ma J, Lenton A. 2001. Imagining stereotypes away: the moderation of automatic stereotypes through mental imagery. *J. Pers. Soc. Psychol.* 81:828–41

Blascovich J, Mendes WB, Hunter SB, Lickel B, Kowai-Bell N. 2001. Perceiver threat in social interactions with stigmatized others. *J. Pers. Soc. Psychol.* 80:253–67

Bosson JK, Swann W, Pennebaker JW. 2000. Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *J. Pers. Soc. Psychol.* 79:631–43

Brauer M, Wasel W, Niedenthal P. 2000. Implicit and explicit components of prejudice. *Rev. Gen. Psychol.* 4:79–101

Brendl CM, Markman AB, Messner C. 2001. How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *J. Pers. Soc. Psychol.* 81:760–73

Bruner JS. 1957. On perceptual readiness. *Psychol. Rev.* 64:123–52

Buchner A, Wippich W. 2000. On the reliability of implicit and explicit memory measures. *Cogn. Psychol.* 40:227–59

Cacioppo JT, Crites SL, Berntson GG, Coles MG. 1993. If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychol. Sci.* 4:108–12

Carlston DE, Smith ER. 1996. Principles of mental representation. See Higgins & Kruglanski 1996, pp. 184–210

Chaiken S, Trope Y, eds. 1999. *Dual Process Theories in Social Psychology*. New York: Guilford

Crites SL, Cacioppo JT, Gardner WL, Bernston GG. 1995. Bioelectrical echoes from evaluative categorization. II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *J. Pers. Soc. Psychol.* 68:997–1013

Cunningham WA, Preacher KJ, Banaji MR. 2001. Implicit attitude measures: consistency, stability, and convergent validity. *Psychol. Sci.* 12:163–70

Dasgupta N, Greenwald AG. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81:800–14

Dasgupta N, McGhee DE, Greenwald AG, Banaji MR. 2000. Automatic preference for white Americans: eliminating the familiarity explanation. *J. Exp. Soc. Psychol.* 36:316–28

De Houwer J. 2001. A structural and process analysis of the Implicit Association Test. *J. Exp. Soc. Psychol.* 37:443–51

De Houwer J. 2002. A structural analysis of indirect measures of attitudes. See Musch & Klauer 2002. In press

De Houwer J, Crombez G, Baeyens F, Hermans D. 2001a. On the generality of the affective Simon effect. *Cogn. Emot.* 15:189–206

De Houwer J, Eelen P. 1998. An affective

variant of the Simon paradigm. *Cogn. Emot.* 12:45–61

De Houwer J, Hermas D, eds. 2001. Automatic affective processing. *Cogn. Emot.* 15:113–248

De Houwer J, Hermans D, Spruyt A. 2001b. Affective priming of pronunciation responses: effects of target degradation. *J. Exp. Soc. Psychol.* 37:85–91

Devine PG, ed. 2001. Implicit prejudice and stereotyping: how automatic are they? *J. Pers. Soc. Psychol.* 71:757–868

Devine PG, Plant EA, Amodio DM, Harmon-Jones E, Vance SL. 2002. The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *J. Pers. Soc. Psychol.* 82:835–48

Dijksterhuis A, Aarts H, Bargh JA, van Knippenberg A. 2000. On the relation between associative strength and automatic behavior. *J. Exp. Soc. Psychol.* 36:531–44

Dovidio JF, Fazio RH. 1992. New technologies for the direct and indirect assessment of attitudes. In *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, ed. JM Tanur, pp. 204–37. New York: Sage

Dovidio JF, Kawakami K, Beach KR. 2001. Implicit and explicit attitudes: examination of the relationship between measures of intergroup bias. In *Blackwell Handbook of Social Psychology: Intergroup Processes*, ed. R Brown, SL Gaertner, pp 175–97. Malden, MA: Blackwell

Dovidio JF, Kawakami K, Gaertner SL. 2002. Implicit and explicit prejudice and interracial interactions. *J. Pers. Soc. Psychol.* 82:62–68

Dovidio JF, Kawakami K, Johnson C, Johnson B, Howard A. 1997. On the nature of prejudice: automatic and controlled processes. *J. Exp. Soc. Psychol.* 33:510–40

Draine SC, Greenwald AG. 1998. Replicable unconscious semantic priming. *J. Exp. Psychol.: Gen.* 127:286–303

Dunton BC, Fazio RH. 1997. An individual difference measure of motivation to control prejudiced reactions. *Pers. Soc. Psychol. Bull.* 23:316–26

Eagly AH, Chaiken S. 1993. *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich

Fazio RH. 1986. How do attitudes guide behavior? In *The Handbook of Motivation and Cognition: Foundations of Social Behavior*, ed. RM Sorrentino, ET Higgins, pp. 204–43. New York: Guilford

Fazio RH. 1990. Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. See Zanna 1990, 23:75–109

Fazio RH. 2001. On the automatic activation of associated evaluations: an overview. *Cogn. Emot.* 15:115–41

Fazio RH, Dunton BC. 1997. Categorization by race: the impact of automatic and controlled components of racial prejudice. *J. Exp. Soc. Psychol.* 33:451–70

Fazio RH, Hilden LE. 2001. Emotional reactions to a seemingly prejudiced response: the role of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Pers. Soc. Psychol. Bull.* 27:538–49

Fazio RH, Jackson JR, Dunton BC, Williams CJ. 1995. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* 69:1013–27

Fazio RH, Sanbonmatsu DM, Powell MC, Kardes FR. 1986. On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50: 229–38

Fazio RH, Towles-Schwen T. 1999. The MODE model of attitude-behavior processes. In *Dual Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 97–116. New York: Guilford

Florack A, Scarabis M, Bless H. 2001. When do associations matter? The use of implicit associations toward ethnic groups in person judgments. *J. Exp. Soc. Psychol.* 37:518–24

Gaertner SL, McLaughlin JP. 1983. Racial stereotypes: associations and ascriptions of positive and negative characteristics. *Soc. Psychol. Q.* 46:23–30

Gawronksi B, Ehrenberg K, Banse R, Zukova J, Klauer KC. 2002. It's in the mind of the

beholder: the impact of stereotypic associations on category-based and individuating impression formation. *J. Exp. Soc. Psychol.* In press

Gilbert DT, Hixon JT. 1991. The trouble with thinking: activation and application of stereotypic beliefs. *J. Pers. Soc. Psychol.* 60:509–17

Greenwald AG, Banaji MR. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102:4–27

Greenwald AG, Banaji MR, Rudman LA, Farnham SD, Nosek BA, et al. 2002a. A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109:3–25

Greenwald AG, Farnham SD. 2000. Using the Implicit Association Test to measure self-esteem and self-concept. *J. Pers. Soc. Psychol.* 79:1022–38

Greenwald AG, Klinger MR, Liu TJ. 1989. Unconscious processing of dichoptically masked words. *Mem. Cogn.* 17:35–47

Greenwald AG, McGhee DE, Schwartz JLK. 1998. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74:1464–80

Greenwald AG, Nosek BA. 2001. Health of the Implicit Association Test at age 3. *Z. Exp. Psychol.* 48:85–93

Greenwald AG, Pickrell JE, Farnham SD. 2002b. Implicit partisanship: taking sides for no reason. *J. Pers. Soc. Psychol.* 83:367–79

Hart AJ, Whalen PJ, Shin LM, McInerney SC, Fischer H, Rauch SL. 2000. Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neurorep.: Rapid Commun. Neurosci. Res.* 11:2351–55

Hastie R. 1984. Causes and effects of causal attribution. *J. Pers. Soc. Psychol.* 46:44–56

Hense RL, Penner LA, Nelson DL. 1995. Implicit memory for age stereotypes. *Soc. Cogn.* 13:399–415

Hermans D, Baeyens F, Eelen P. 2002. On the acquisition of evaluative information in memory: the study of evaluative learning and affective priming combined. See Musch & Klauer 2002. In press

Hetts JJ, Sakuma M, Pelham BW. 1999. Two roads to positive regard: implicit and explicit self–evaluation and culture. *J. Exp. Soc. Psychol.* 35:512–59

Higgins ET. 1996. Knowledge activation: accessibility, applicability, and salience. See Higgins & Kruglanski 1996, pp. 133–68

Higgins ET, Kruglanski AW, eds. 1996. *Social Psychology: Handbook of Basic Principles*. New York: Guilford

Ito TA, Cacioppo JT. 2000. Electrophysiological evidence of implicit and explicit categorization processes. *J. Exp. Soc. Psychol.* 36: 660–76

Jackson JR. 1997. *Automatically activated racial attitudes*. PhD thesis, Indiana Univ.

Jones JT, Pelham BW, Mirenberg MC, Hetts JJ. 2002. Name letter preferences are not merely mere exposure: implicit egotism as self-regulation. *J. Exp. Soc. Psychol.* 38:170–77

Jordan CH, Spencer SJ, Zanna MP. 2002. I love me . . . I love me not: implicit self-esteem, explicit self-esteem, and defensiveness. In *Motivated Social Perception: The Ninth Ontario Symposium*, ed. SJ Spencer, S Fein, MP Zanna, JM Olson, pp. 117–45. Mahwah, NJ: Erlbaum

Jost JT, Pelham BW, Carvallo MR. 2002. Non-conscious forms of system justification: implicit and behavioral preferences for higher status groups. *J. Exp. Soc. Psychol.* In press

Karpinski A, Hilton JL. 2001. Attitudes and the Implicit Association Test. *J. Pers. Soc. Psychol.* 81:774–78

Kawakami K, Dion KL, Dovidio JF. 1998. Racial prejudice and stereotype activation. *Pers. Soc. Psychol. Bull.* 24:407–16

Kawakami K, Dovidio JF. 2001. The reliability of implicit stereotyping? Or confronting prejudice: effects of prejudice labeling on stereotype activation. *Pers. Soc. Psychol. Bull.* 27: 212–25

Kawakami K, Dovidio JF, Moll J, Hermsen S, Russin A. 2000. Just say no to stereotyping: effects of training in the negation of

stereotypic associations on stereotype activation. *J. Pers. Soc. Psychol.* 78:871–88

Keuhnen U, Schiessl M, Bauer M, Paulig N, Poehlmann C, et al. 2001. How robust is the IAT? Measuring and manipulating attitudes of East and West-Germans. *Z. Exp. Psychol.* 48:135–44

Klauer KC. 1998. Affective priming. *Eur. Rev. Soc. Psychol.* 8:63–107

Klinger MR, Burton PC, Pitts GS. 2000. Mechanisms of unconscious priming. I. Response competition, not spreading activation. *J. Exp. Psychol.: Learn. Mem. Cogn.* 26:441–55

Koole SL, Dijksterhuis A, van Knippenberg A. 2001. What's in a name: implicit self-esteem and the automatic self. *J. Pers. Soc. Psychol.* 80:669–85

Lemm KM. 2001. *Personal and social motivation to respond without prejudice: implications for implicit and explicit attitude and behavior*. PhD thesis, Yale Univ.

Lepore L, Brown R. 1997. Category and stereotype activation: Is prejudice inevitable? *J. Pers. Soc. Psychol.* 72:275–87

Livingston RW. 2002. The role of perceived negativity in the moderation of African-Americans' implicit and explicit racial attitudes. *J. Exp. Soc. Psychol.* 38:405–13

Livingston RW, Brewer MB. 2002. What are we really priming? Cue-based versus category-based processing of facial stimuli. *J. Pers. Soc. Psychol.* 82:5–18

Lord CG, Lepper MR. 1999. Attitude representation theory. In *Advances in Experimental Social Psychology*, ed. MP Zanna, 31:265–343. San Diego, CA: Academic

Lord CG, Lepper MR, Mackie D. 1984. Attitude prototypes as determinants of attitude-behavior consistency. *J. Pers. Soc. Psychol.* 46:1254–66

Lowery BS, Hardin CD, Sinclair S. 2001. Social influence effects on automatic racial prejudice. *J. Pers. Soc. Psychol.* 81:842–55

Maass A, Ceccarelli R, Rudin S. 1996. Linguistic intergroup bias: evidence for in-group-protective motivation. *J. Pers. Soc. Psychol.* 71:512–26

Maass A, Salvi L, Arcuri L, Semin GR. 1989. Language use in intergroup contexts: the linguistic intergroup bias. *J. Pers. Soc. Psychol.* 57:981–93

Macrae CN, Bodenhausen GV, Milne AB, Thorn TMJ, Castelli L. 1997. On the activation of social stereotypes: the moderating role of processing objectives. *J. Exp. Soc. Psychol.* 33:471–89

Maison D, Greenwald AG, Bruin R. 2001. The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychol. Bull.* 32:61–69

Marsh KL, Johnson BL, Scott-Sheldon LA. 2001. Heart versus reason in condom use: implicit versus explicit attitudinal predictors of sexual behavior. *Z. Exp. Psychol.* 48:161–75

McConnell AR, Liebold JM. 2001. Relations between the Implicit Association Test, explicit racial attitudes, and discriminatory behavior. *J. Exp. Soc. Psychol.* 37:435–42

Mierke J, Klauer KC. 2001. Implicit association measurement with the IAT: evidence of the effects of executive control processes. *Z. Exp. Psychol.* 48:107–22

Mitchell JA, Nosek BA, Banaji MR. 1999. *Dissociated implicit attitudes: examples from race, gender, and profession.* Presented at Annu. Meet. Midwestern Psychol. Assoc., Chicago

Monteith MJ, Voils CI, Ashburn-Nardo L. 2001. Taking a look underground: detecting, interpreting, and reacting to implicit racial bias. *Soc. Cogn.* 19:395–417

Musch J, Klauer KC, eds. 2002. *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. Mahwah, NJ: Erlbaum. In press

Nosek BA. 2002. *Moderators of the relationship between implicit and explicit attitudes.* PhD thesis, Yale Univ.

Nosek BA, Banaji MR. 2001. The go/no-go association task. *Soc. Cogn.* 19:625–66

Nosek BA, Banaji MR, Greenwald AG. 2002a. Math = Male, Me = Female, therefore Math is not equal to Me. *J. Pers. Soc. Psychol.* 83:44–59

Nosek BA, Banaji MR, Greenwald AG. 2002b.

Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynam.* 6:101–15

Nuttin JM. 1985. Narcissism beyond Gestalt and awareness: the name letter effect. *Eur. J. Soc. Psychol.* 15:353–61

Olson MA, Fazio RH. 2001. Implicit attitude formation through classical conditioning. *Psychol. Sci.* 12:413–17

Olson MA, Fazio RH. 2002a. Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Basic Appl. Soc. Psychol.* In press

Olson MA, Fazio RH. 2002b. Implicit acquisition and manifestation of classically conditioned attitudes. *Soc. Cogn.* 20:89–104

Ottaway SA, Hayden DC, Oakes MA. 2001. Implicit attitudes and racism: effects of word familiarity and frequency in the Implicit Association Test. *Soc. Cogn.* 19:97–144

Otten S, Wentura D. 1999. About the impact of automaticity in the minimal group paradigm: evidence from affective priming tasks. *Eur. J. Soc. Psychol.* 29:1049–71

Pelham BW, Mirenberg MC, Jones JK. 2002. Why Susie sells seashells by the seashore: implicit egotism and major life decisions. *J. Pers. Soc. Psychol.* 82:469–87

Perdue CW, Dovidio JF, Gurtman MB, Tyler RB. 1990. Us and them: social categorization and the process of intergroup bias. *J. Pers. Soc. Psychol.* 59:475–86

Perruchet P, Baveux P. 1989. Correlational analyses of explicit and implicit memory performance. *Mem. Cogn.* 17:77–86

Phelps EA, O'Connor KJ, Cunningham WA, Funayama ES, Gatenby JC, et al. 2000. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12:729–38

Plessner H, Banse R, eds. 2001. Attitude measurement using the Implicit Association Test (IAT). *Z. Exp. Psychol.* 48:82–175

Pratto F, Shih M. 2000. Social dominance orientation and group context in implicit group prejudice. *Psychol. Sci.* 11:515–18

Proshansky HM. 1943. A projective method for the study of attitudes. *J. Appl. Soc. Psychol.* 38:393–95

Richardson-Klavehn A, Bjork RA. 1988. Measures of memory. *Annu. Rev. Psychol.* 39:475–543

Richeson JA, Ambady N. 2002. Effects of situational power on automatic racial prejudice. *J. Exp. Soc. Psychol.* In press

Roediger HL. 1990. Implicit memory: retention without remembering. *Am. Psychol.* 45: 1043–56

Rothermund K, Wentura D. 2001. Figure-ground asymetries in the Implicit Association Test. *Z. Exp. Psychol.* 48:94–106

Rudman LA, Ashmore RD, Gary ML. 2001a. "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *J. Pers. Soc. Psychol.* 81:856–68

Rudman LA, Glick P. 2001. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* 57:743–62

Rudman LA, Greenwald AG, McGhee DE. 2001b. Implicit self-concept and evaluative implicit gender stereotypes: self and ingroup share desirable traits. *Pers. Soc. Psychol. Bull.* 27:1164–78

Rudman LA, Greenwald AG, Mellott DS, Schwartz JLK. 1999. Measuring the automatic components of prejudice: flexibility and generality of the Implicit Association Test. *Soc. Cogn.* 17:437–65

Rudman LA, Kilianski SE. 2000. Implicit and explicit attitudes toward female authority. *Pers. Soc. Psychol. Bull.* 26:1315–28

Rudman LA, Lee MR. 2002. Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Process. Intergroup Relat.* 5:133–50

Schacter DL. 1987. Implicit memory: history and current status. *J. Exp. Psychol.: Learn. Mem. Cogn.* 13:501–18

Schneider W, Fisk AD. 1982. Degree of consistent training: improvements in search performance and automatic process development. *Percept. Psychophys.* 31:160–68

Schneider W, Shiffrin RM. 1977. Controlled and automatic human information

processing. I. Detection, search, and attention. *Psychol. Rev.* 84:1–66

Schwarz N, Bohner G. 2001. The construction of attitudes. In *Blackwell Handbook of Social Psychology: Intraindividual Processes*, ed. A Tesser, N Schwarz, pp. 436–57. Malden, MA: Blackwell

Sekaquaptewa D, Espinoza P, Thompson M, Vargas P, von Hippel W. 2002. Stereotype explanatory bias: implicit stereotyping as a predictor of discrimination. *J. Exp. Soc. Psychol.* In press

Semin GR, Fiedler K. 1988. The cognitive functions of linguistic categories in describing persons: social cognition and language. *J. Pers. Soc. Psychol.* 54:558–68

Sherif M, Hovland CI. 1961. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven, CT: Yale Univ. Press

Sherman SJ, Presson CC, Chassin L, Rose JS, Koch K. 2002. Implicit and explicit attitudes toward cigarette smoking: the effects of context and motivation. *J. Soc. Clin. Psychol.* In press

Sinclair L, Kunda Z. 1999. Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *J. Pers. Soc. Psychol.* 77:885–904

Smith ER, Fazio RH, Cejka MA. 1996. Accessible attitudes influence categorization of multiply categorizable objects. *J. Pers. Soc. Psychol.* 71:888–98

Snyder M, Kendzierski D. 1982. Acting on one's attitudes: procedures for linking attitude and behavior. *J. Exp. Soc. Psychol.* 18:165–83

Snyder M, Swann WB. 1976. When actions reflect attitudes: the politics of impression management. *J. Pers. Soc. Psychol.* 34:1034–42

Son Hing LS, Li W, Zanna MP. 2002. Inducing hypocrisy to reduce prejudicial responses among aversive racists. *J. Exp. Soc. Psychol.* 38:71–78

Spalding LR, Hardin CD. 1999. Unconscious unease and self-handicapping: behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychol. Sci.* 10:535–39

Stacy AW, Leigh BC, Weingardt K. 1997. An individual difference perspective applied to word association. *Pers. Soc. Psychol. Bull.* 23:229–37

Steffens MC, Plewe I. 2001. Items' cross-category associations as a confounding factor in the Implicit Association Test. *Z. Exp. Psychol.* 48:123–34

Swanson JE, Rudman LA, Greenwald AG. 2001. Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cogn. Emot.* 15:207–30

Teachman BA, Gregg AP, Woody SR. 2001. Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *J. Abnorm. Psychol.* 110:226–35

Towles-Schwen T, Fazio RH. 2002. Choosing social situations: the relation between automatically-activated racial attitudes and anticipated comfort interacting with African Americans. *Pers. Soc. Psychol. Bull.* In press

Tulving E, Schacter DL, Stark HA. 1982. Priming effects in word-fragment completion are independent of recognition memory. *J. Exp. Psychol.: Learn. Mem. Cogn.* 8:336–42

Vanman EJ, Paul BY, Ito TA, Miller N. 1997. The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *J. Pers. Soc. Psychol.* 73:941–59

von Hippel W, Sekaquaptewa D, Vargas P. 1995. On the role of encoding processes in stereotype maintenance. In *Advances in Experimental Social Psychology*, ed. MP Zanna, 27:174–255. San Diego, CA: Academic

von Hippel W, Sekaquaptewa D, Vargas P. 1997. The linguistic intergroup bias as an implicit indicator of prejudice. *J. Exp. Soc. Psychol.* 33:490–509

Warrington EK, Weiskrantz L. 1968. New method of testing long-term retention with special reference to amnesic patients. *Nature* 217:972–74

Webb EJ, Campbell DT, Schwartz RD, Sechrest L. 1966. *Unobtrusive Measures: Nonreactive*

Research in the Social Sciences. Chicago: Rand McNally

Wegener DT, Petty RE. 1995. Flexible correction processes in social judgment: the role of naive theories in corrections for perceived bias. *J. Pers. Soc. Psychol.* 68:36–51

Wentura D. 1999. Activation and inhibition of affective information: evidence for negative priming in the evaluation task. *Cogn. Emot.* 13:65–91

Wilson TD, Lindsey S, Schooler TY. 2000. A model of dual attitudes. *Psychol. Rev.* 107:101–26

Wittenbrink B, Judd CM, Park B. 1997. Evidence for racial prejudice at the implicit level and its relationship to questionnaire measures. *J. Pers. Soc. Psychol.* 72:262–74

Wittenbrink B, Judd CM, Park B. 2001a. Spontaneous prejudice in context: variability in automatically activated attitudes. *J. Pers. Soc. Psychol.* 81:815–27

Wittenbrink B, Judd CM, Park B. 2001b. Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *J. Exp. Soc. Psychol.* 37:244–52

Zanna MP, Fazio RH. 1982. The attitude-behavior relation: moving toward a third generation of research. In *Consistency in Social Behavior: The Ontario Symposium*, ed. MP Zanna, ET Higgins, CP Herman, 2:283–301. Mahwah, NJ: Erlbaum

Zanna MP, Rempel JK, 1988. Attitudes: a new look at an old concept. In *The Social Psychology of Knowledge*, ed. D Bar-Tal, AW Kruglanski, pp. 315–34. New York: Cambridge Univ. Press

Annual Review of Psychology
Volume 54, 2003

CONTENTS