

In Roediger, H.L., Nairne, J.N., Suprenant, A.M. (2001). The nature of remembering: Essays in honor of Robert G. Crowder. Washington, D.C., APA, 117-149.

CHAPTER 7

Implicit Attitudes Can Be Measured

Mahzarin R. Banaji

The title of this essay is bold and unoriginal. It is, as those of advanced age or schooling will recognize, a variation of L. L. Thurstone's (1928) landmark article "Attitudes Can Be Measured," in which he described his general method of equal appearing intervals, introducing the topic thus:

The natural first impression about these two concepts [attitude and opinion] is that they are not amenable to measurement in any real sense. It will be conceded at the outset that an attitude is a complex affair which cannot be wholly described by any single numerical index. For the problem of measurement this statement is analogous to the observation that an ordinary table is a complex affair which cannot be wholly described by any single numerical index. So is a man such a complexity

This chapter was difficult to write, in part, because of the nature of the subject matter to be tackled and, in part, because it was a painful reminder of Bob Crowder's death. He will never get to read it, but I know exactly the parts that would have made him smile.

Not only did R. Bhaskar and Richard Hackman encourage that these ideas be put into print (if only to stop my increasing groaning about them), but their presence is palpable in every paragraph, except this one. Tony Greenwald brilliantly produced the Implicit Association Test (IAT); need I say more. Claude Steele provided affirmation that the IAT effect was the phenomenon of interest, independent of immediate prediction. Many colleagues provided comments on short notice: Roddy Roediger made observations and corrections without which the scholarship would be poorer by several just noticeable differences. Paul Bloom, Siri Carpenter, Wil Cunningham, Buju Dasgupta, Thierry Devos, Stephanie Goodwin, Jack Glaser, Tony Greenwald, Larry Jacoby, Kristin Lane, Kristi Lemm, Brian Nosek, Jason Mitchell, Wayne Steward, Laurie Rudman, and Bethany Teachman provided invaluable comments.

For many years, Susan Bouregy and Suzanne Polmar of Yale University's Faculty of Arts and Sciences Committee on Research on Human Subjects have provided unflagging support, in spite of the disproportionate amount of "business" my group sends their way. National Science Foundation (NSF) Grants SBR 9422241 and 9709924 and National Institute of Mental Health (NIMH) Grant MH-57672 supported the research and writing. I am grateful beyond dollars. Especially, I thank Carolyn Morf of NIMH and Steven Breckler of NSF for their vigilant monitoring. These grants compensate for my working at a university with a small course-based respondent pool and provide support for the finest graduate students there are (who dared to come here anyway).

which cannot be wholly represented by a single index. Nevertheless we do not hesitate to say that we measure the table. The context usually implies what it is about the table that we propose to measure. We say without hesitation that we measure a man when we take some anthropometric measurements of him . . . his height or weight or what not. Just in the same sense we shall say here that we are measuring attitudes. . . . The point is that it is just as legitimate to say that we are measuring attitudes as it is to say that we are measuring tables or men. (pp. 530–531)

With these words, Thurstone launched into the specifics of his idea that evenly graduated opinions could be so arranged that equal steps or intervals on the attitude scale represented equally noticeable shifts in attitude. Of interest to me here is not the particular method Thurstone devised but rather the radical assessment offered by his generation of social psychologists—the idea that the fundamental act of evaluation, of mentally appraising an object along a good–bad dimension, could itself be measured. Until then, attitude and associated constructs such as feeling, evaluation, and preference were considered to be sufficiently intractable as to elude capture in the laboratory. This, of course, explains Thurstone's insistence that attitudes could indeed be measured, even though the need for such exhortation must surely seem quaint to early-21st-century psychologists.

It would appear that research on attitudes, established in 1935 as social psychology's "most central and indispensable construct" (Allport, 1954), should provide sufficient preparation for future developments of both theory and measurement. It would also appear that the secure early groundwork for attitude measurement should provide a sturdy enough foundation to explore even unorthodox possibilities for this cornerstone construct. Yet surprisingly, that does not seem to be the case as the field confronts new, previously undetected forms of attitudes. Persistent questions arise about the measurement of certain attitudes, especially when they appear not in their familiar, conscious form but as they materialize in ways that elude conscious awareness, seem oblivious to conscious intention, and defy conscious control.

I should assure readers that the troubles that such attitudes and their measurement elicit are not imaginary issues created largely for my own delectation. Although not yet available in print, they have been persistently posed and with a passion atypical of the sterility of normal academic exchange. Large numbers of individuals, both academic colleagues and lay audiences, have raised questions in the course of many presentations (over 100 conference and colloquia presentations in the last 5 years by members of the lab), in the course of submissions to the Yale University Psychology Department's Committee on Research on Human Subjects and in the form of written communications. Their interest in this work has been marked by an admirable curiosity and the deepest insight about a phenomenon that can, at times, be disconcerting. I am deeply obliged to the individuals whose comments improved the research and directly influenced this report. That said, this chapter contains responses to what I regard to be an orthodoxy that needs to be challenged if a new generation is to probe in more intrepid fashion the existence of implicit attitudes and their measurement.

Interestingly, in this particular line of work, issues and questions raised by experts and lay audiences appear to be equal in sophistication. Here, I raise primarily those issues that have a broad consensus across both communities, with obvious deference to issues of scientific rather than social or political import. The particular implicit measures of attitude with which I am concerned consist of identifying an attitude object (which could be anything: the NASDAQ index, single malt scotch, Timothy McVeigh, the African National Congress, or god) and pairing it in a variety of ways with evaluative attributes (i.e., the concepts of pleasant or unpleasant, good or bad). The speed of responding to say "god" when it is paired with "pleasant" than "unpleasant" is taken as a measure of the individual's strength of positivity or negativity toward the concept "god." A great deal of work has now been produced that uses variations of measures of automatically activated attitudes (see Banaji, Lemm, & Carpenter, in press; Fazio, in press; Greenwald & Banaji, 1995).

Because the theme of this book is memory, I would have liked to create a two-way street with equal traffic along attitude and memory lanes. However, in this chapter, I necessarily emphasize implicit attitudes and beliefs with comparisons to unconscious memory. Invisibly beneath the analysis lies an assessment offered by Gordon Allport (1954) in his famous tract, *The Nature of Prejudice*—namely, that the concepts of attitude (prejudice) and belief (stereotype) are deeply rooted in the ordinary ways in which humans perceive, categorize, and remember (see Banaji & Bhaskar, 1999, for implications of this view for the bounded rationality of social judgment). In addition, because the analyses presented here necessitate connecting constructs that arbitrarily happen to sit in distinct academic spheres (i.e., attitude–belief in social psychology and memory in cognitive psychology), some of what I say may seem either too obvious or too unfamiliar to readers with differing intellectual alliances. I take this risk happily, in part because of a comment Bob Crowder once made about my uneasy perch at the intersection of areas—"but we are both interested in how people think," he said—a comment that was as comforting more than a dozen years ago as it is now, when the demand for an integrated analysis of mind and society appears to have raised new possibilities.

In this chapter, I point out difficulties with the concept of implicit attitude and possible reasons for them. The discussion focuses on the two main questions that arise most persistently in response to the use of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), and the goal is to provide interim answers to each. The first question may be boiled down to this: Are implicit attitudes really attitudes at all? More to the point, are presumed measures of implicit attitudes (i.e., measures of the strength of association between evaluation and attitude object) detecting attitudes at all? Second, how valid are implicit attitudes? In particular, what are they related to and distinct from, and what do they predict in the so-called real world? I address these issues here in part because my thinking about implicit processes (not to mention the real world) was shaped by a dozen years of interaction with Bob Crowder, most directly through his weekly "memory lunch" meetings.

Less nostalgically, I see the constructs of attitude and memory as deeply intertwined, and I am recklessly optimistic that any assessments of implicit attitudes that are misguided, including my own, will fade when considered in the context of parallel developments in the study of implicit memory. Conversely, questions concerning the reliability of implicit measures of memory and the relationship among such measures that are sometimes swept under the rug may be clarified in the light of discoveries about implicit evaluation and attitude.

Are Implicit Attitudes Attitudes?

Construals of Implicit Memory

There is incomplete agreement regarding the use of the term "implicit memory" ("unconscious memory" and "indirect measure of memory" are other contenders). I nevertheless use it here because in my own work, I have made use of that term when applied to social cognition (see, e.g., Banaji, Hardin, & Rothman, 1993; Greenwald & Banaji, 1995). I begin with the construct of implicit memory and the family of tasks that constitute its main measures. As students of memory know, the concept of unconscious or implicit memory came into prominence in the late 1970s and early 1980s because of an exciting convergence of evidence from patient and college samples using both clinical observations and experimental methods. This body of research revealed the existence of new forms of memory that lay well concealed from conscious recollection yet were starkly present on indirect measures, such as lexical decisions, perceptual identification, and other seemingly irrelevant judgments (Jacoby, 1994; Richardson-Klavhen & Bjork, 1988; Roediger, 1990; Roediger & McDermott, 1993; Schacter, 1987). Such research followed on the heels of a previous decade of work (Meyer & Schvaneveldt, 1971; Neely, 1977), active even today, that used a lexical decision judgment to gauge automatic processes in semantic memory. Taken together, such research made the psychological study of two central facets of consciousness immensely more tractable: the study of memory as it operated outside conscious awareness and memory as it operated without conscious control.

To investigate alternative forms of memory, one needed new measures, and a remarkably ingenious set came to be devised within a short span of time to allow discoveries of dissociated memory processes or memory systems. In addition to the heavily used lexical decision task, other tasks such as perceptual identification, word-stem completion, word- and picture-fragment naming, and a host of judgment tasks (e.g., fame of a name, liking for an object, brightness and loudness estimates) served as indirect measures of implicit memory, with the signature finding being a dissociation between conscious and unconscious forms.

In regard to the results from such experiments, it should not be overlooked that a paradigm shift had occurred. Measurement and theory worked hand-in-hand

to produce breakthroughs that transformed the understanding of the nature of memory. Memory, which for 100 years had been generally equated with recollection (as in the act of reflecting consciously on things past), became stretched to accommodate a broader meaning. This expanded meaning of memory, at least superficially, bore faint resemblance to the older conception of what memory meant to laypeople and to scientists. Nowhere was the difference more striking than in the measurement tools themselves, which no longer required conscious recollection on the part of the participant. Compared with traditional, face-valid measures of free recall and recognition, measures of implicit memory could, at least to the naked eye, seem alien. The comfortable assumption, that everybody and their uncle Chet could accept, was that if memory for the past is required, well then, just ask. Implicit measures, however, require imagining the improbable: For example, one takes the speed (or ability) to respond to a quickly flashing word, or the proficient completing of a word fragment, as an indicator of memory. Or even more absurd to the unprepared mind, one takes subjective assessments of a stimulus, such as the perceived fame of a name or expressed liking for a polygon or a human face, as a measure of memory.

Difficult as such leaps may have been, these seemingly alien measures were seen to reveal something new about memory. The important point is that the novel nature of these measures did not lead investigators to scramble around for a new construct or to become baffled about which was the "real" memory (the one that responded with sensitivity to free recall, for instance, or the one that showed no budging on free recall but an almost aesthetic sensitivity to perceptual identification).¹ Seeing that the old and the new were possibly part and parcel of the same construct of memory seemed possible not so much on the basis of grand theory but rather as a bottom-up attempt to make sense of intriguing phenomena—about dissociated memory processes in people with ordinary memory and in the stark observations of amnesia patients that implied a deep separation of memory systems. In any event, the same construct of memory came to be expanded to now include varyingly named extensions, such as unconscious memory, implicit memory, automatic memory, indirect memory, and procedural memory.

¹Pioneers of research on unconscious forms of memory have pointed out there was resistance to the concept of unconscious memory when it was first introduced that may still linger. Yet one has to assume that such reactions were marginalized because research on implicit memory is thriving and has been an important player in the new mind sciences. Both Roddy Roediger and Larry Jacoby (personal communications, January 26–27, 2000) noted that there were indeed those who did not accept that measures of implicit memory revealed memory at all. A reviewer of Jacoby and Dallas (1981), for instance, thought the article interesting but not about memory; the article is now widely regarded as an important contribution to the understanding of memory.

Construals of Implicit Attitudes

It is awkward to discuss metamethodological issues concerning implicit attitudes and their parallels to memory without a clear sense of just what is meant by implicit attitudes. I therefore move to a discussion of definitions and the properties of particular measures. To allay fears that implicit attitudes are not really attitudes, Jason Mitchell (personal communications, May 2000) used an amusing strategy of plotting an array of three definitions in his lectures. His goal was to show the correspondence between the two firmly established definitions of attitude and implicit memory and the resulting definition of implicit attitudes. If attitude is "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 1998, p. 269) and if implicit memory is "revealed when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of these experiences" (Schacter, 1987, p. 501), then implicit attitudes are "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald & Banaji, 1995, p. 8).

At the level of abstract concepts, as reflected in the definitions, there is reassuring congruity. But this, as seen later, is not at issue. Difficulties begin when particular measures do not sit well with established notions of how the attitude construct is being conceptualized now, in contrast to the past, or to its lay meaning. These aspects of measurement, rather than any definitional imbalances, appear to be the source of concern.

The family of implicit attitude measures is becoming increasingly diverse (see Fazio, in press), a positive state of affairs insofar as it does not result in "camps" on the basis of task rather than theory. I describe two variations, evaluative priming and the IAT, although this family of tasks, including the affective Stroop task, the affective flanker task, and other response competition tasks, contains many similarities and surprising dissimilarities (De Houwer, 1999).

Evaluative Priming as a Measure of Attitude

Perhaps the greatest amount of published research on automatic attitudes to date uses an evaluative priming task, brought to prominence by Russ Fazio and his colleagues (see, e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The evaluative priming task closely follows the procedure used to measure semantic priming (see Fazio et al., 1986; see also Bargh, Chaiken, Govender, & Pratto, 1992; and Glaser & Banaji, 1999). In evaluative priming, primes (attitude objects) are followed by targets (evaluative words), and the speed to judge the target is taken as an indicator of the strength of evaluative association between attitude object and evaluation. In essence, the assumption is that if an attitude object (e.g., flowers, ice cream) evokes a positive evaluation, it should facilitate response to other evaluatively congruent (i.e., positive) co-occurring information; likewise, if an attitude object (e.g., insects,

anchovies) evokes a negative evaluation, it ought to be easier to respond to evaluatively negative co-occurring material.

In an evaluative priming task, a prime containing the attitude object (e.g., flower, insect, ice cream, anchovies) is presented for a short duration (say 200 ms) followed by an evaluatively congruent or incongruent but semantically unassociated target (e.g., friend, agony). Responses of "good or bad," "word or nonword," or mere pronunciation of the target are facilitated to the extent to which prime and target are evaluatively related. This evaluative priming effect has been taken to be a measure of automatic attitude. Such findings, and even their interpretation, appear to be well accepted—certainly my own research using such a procedure both in the domain of attitudes (Glaser & Banaji, 1999) and implicit beliefs or stereotypes (Banaji & Hardin, 1996; Blair & Banaji, 1996) has been interpreted as such—measures of implicit attitude or stereotype. It is important to note that a further interpretation of the strength of evaluative associations to social groups, Black and White, reveals the strength of anti-Black prejudice, has also been offered (Fazio, Jackson, Dunton, & Williams, 1995), and preliminary evidence of its validity affirmed through the prediction of a nonverbal measure of prejudicial behavior. That is, implicit pro-White attitude (prejudice), as measured by the strength of evaluative association between Black + bad, White + bad (and likewise of association with "good" as well), predicted the degree of anti-Black, nonverbal action.

The Implicit Association Test

Since 1995, another measure of automatic attitudes, the IAT, has been in use, although the first article on it appeared much later (Greenwald et al., 1998) to ensure satisfactory parameter testing. The IAT relies on a response latency indicator obtained in the process of pairing the attitude object (e.g., a social group such as old-young) with an evaluative dimension (e.g., good-bad) or specific attributes that may not be purely evaluative (e.g., self-other, home-career, science-arts). In the computerized version, the pairing is achieved by using a keyboard key (e.g., a left key) to be pressed in response to items from the two paired categories (e.g., old + bad) while another key (e.g., a right key) is used for the other pair (e.g., young + good). The speed to complete this pairing compared with the opposite one is interpreted as a measure of the strength of implicit evaluation (attitude). Likewise, the strength of association between concept + attribute (i.e., old-young with home-career) is interpreted to be a measure of the strength of implicit belief (stereotype). In the case of the IAT, association between attitude object and evaluation is achieved by pairing many instances of the attitude object (e.g., flower: daffodil, rose, tulip) with instances of a particular evaluative category (e.g., good: vacation, cake, truth) through arbitrary assignment to a response key (left or right). For example, if implicit attitudes toward flowers are to be measured, a contrasting category is generated (e.g., insects) and the task requires all instances of the category flower

and the category good to be assigned to a single (e.g., left) response key, while all instances of the category insect and the opposing evaluative category bad are assigned to a different (e.g., right) response key. Responses in the condition (flower + good and insect + bad) are compared with responses made in the opposite pairing (flower + bad and insect + good). The difference in response latencies to items in these two conditions of alternating pairings is taken as the measure of implicit attitude toward flowers versus insects. The IAT effect is a difference score reflecting a relative attitude that shows both the direction (positive vs. negative) of implicit attitude and the magnitude of the attitude (larger numbers reflecting larger differences between pairings in milliseconds).

Evaluative priming and IAT tasks share several basic assumptions: (a) that strength of evaluative association can be measured; (b) that the extent to which concepts share evaluative meaning (independent of semantic meaning) is revealed in the ease with which they can be mentally paired; (c) that one way to measure the strength of evaluative association is to measure the speed of object + evaluation pairs; and (d) that the strength of evaluative association as measured under conditions of speeded responding is a measure of automatic attitude. Most important, both tasks measure the strength of evaluative association in some way, and both take the strength of that evaluative association to reflect the strength of automatic attitude—that is their fundamental commonality. To be sure, there are many obvious and subtle differences between these methods, and among the goals for future research lies the task of understanding the processes that each taps and its unique prediction of the downstream effects of social judgment.

So Are Implicit Attitudes Attitudes?

Having offered definitions of implicit attitudes and provided a summary of two measures, I redirect the discussion to the question of the meaning of implicit attitudes. One issue concerns just how tightly an experimental operation or specific task should be tied to the psychological construct assumed to be underlying it. Although semantic priming provided a measure of the strength of association between prime and target pairs, the findings were not interpreted as indicative merely of measures of associative strength alone (although surely that was the implied mechanism) but rather as measures of semantic memory. Likewise, data from word-stem completion tasks were not regarded as speaking to the question of accessibility alone (although surely that is a reasonable description of what prior exposure does) but rather as evidence of savings in memory. Whatever may have been the informal discussions among investigators at the time the paradigm was shifting, the outcome, as known today, is this: Although new measures and emerging new conceptualizations of memory bore no family resemblance to their ancestors, they were regarded as speaking to the question of memory, not some interim process such as accessibility or associative strength alone. The extension to implicit attitudes is simple: I would

claim that it is attitude, as defined conventionally, that these new measures speak to, albeit in a form that differs from the conventional one. The lack of intuitive appeal of measures of implicit attitude should not detract from the expansion of the attitude construct to include such measures.

In selecting an example target to demonstrate the difficulty of imagining new measures of old constructs, I have taken the conventional South Indian position that those easiest to exploit to support one's parochial position are those who are closest, that is, one's own family. Here, I shamelessly use a group to which I am tightly bound by ties of affection including collaboration, car rides, and babysitting—a segment of Yale's Psychology Department that constitutes a committee on which I have myself served many a time. The following local example is therefore used to illustrate the mental hurdle I describe concretely.

Psychologists at Yale University may encounter two committees on ethics concerning human research: a Faculty of Arts and Sciences Committee and a second, departmental committee, whose role is to determine whether the proposed research additionally meets an educational goal if participants are to be drawn from Introductory Psychology courses. Although my research has always been routinely approved by both committees, it has elicited a peculiar scrutiny from the departmental committee—along a dimension that some may even regard to be outside the purview of the committee's task. That is, the departmental committee has chosen to specify the theoretical interpretation that participants may receive during debriefing and the theoretical interpretation that is likewise prohibited. I quote from two letters written recently by this committee: "The IAT should be described as a measure of implicit associations rather than a measure of implicit attitudes, beliefs, preferences, or stereotypes" (a letter to Brian Nosek, a student colleague of mine, dated January 23, 2000), and similarly, "the first concern [of this committee] has to do with your reference to the IAT, as a measure of attitudes (a claim some Committee members are uncomfortable with) instead of a measure of implicit associations (which none are uncomfortable with)" (a letter to Kristin Lane, a student colleague of mine, dated January 23, 2000).

I should be clear that the views of this committee are widely shared among its members and have been consistently expressed over time. As a result, I take these assessments with the utmost seriousness, in the sense that I adhere to the prescriptions of the committee. Nevertheless, I remain unpersuaded by its assessment, about as unpersuaded as somebody studying memory using a lexical decision task might be if told that "the first concern of this committee has to do with your reference to the lexical decision task as a measure of memory (a claim some committee members are uncomfortable with) instead of a measure of automatic association (which none are uncomfortable with)." The fact that such an issue arises suggests a genuine inability to accommodate a new manifestation of even a familiar and well-worn concept.

The nature of debriefing here offers an opportunity to examine the different standards that are set when a procedure is unfamiliar. It is, of course, important in any

education of participants in a research enterprise to take seriously the interpretation of the specific procedures of the experiment offered. In the case of the experiments in question, participants should be (and are) made aware that what was measured is the speed to respond to concept terms (e.g., old–young) when paired with Attribute A (e.g., good) relative to the speed when paired with (a typically complementary) Attribute B (e.g., bad). Furthermore, participants should also be made aware that this speed to respond to particular pairings of stimuli has been viewed, over decades of research, to be indicative of the strength of association between concepts (e.g., between doctor–nurse vs. doctor–butcher).

To leave it at that, however easy it appears, seems dissatisfying, not to mention the ethical dilemma it poses. In the spirit of full debriefing, there is an obligation to offer participants the understanding available at present that implicit evaluations have been regarded, by experts, to reveal attitude since at least 1986 (Fazio et al., 1986). Also such an interpretation remains within the bounds of widely endorsed definitions of attitudes, as a “psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly & Chaiken, 1998, p. 269). On the basis of 2 decades of research (see Bargh et al., 1992; Fazio, Chen, McDonel, & Sherman, 1982; Fazio et al., 1986; Glaser & Banaji, 1999) in which the automatic association between a concept and an evaluative attribute has routinely been discussed as an attitude, should one not be able to speak about the construct using that term? These questions are not the banal ones they may seem about the language of debriefing in a chapter on the measurement of implicit attitudes. Instead, they get to the heart of how one may responsibly go about determining the construct validity of these measures.

A Personal Perspective

When I first came to consider the idea of unconscious social cognition, it was not without difficulty that I cut loose from the dominant view of attitudes as conscious and therefore as open to change by sheer dint of will. My mind was changed through the insights and discoveries of others, primarily by the two sources I mention here. I report them not because I wish to drag readers through the Sisyphean development of my own thinking but because I regard one of these sources to be of historical importance in the development of the concept of automatic attitude (Fazio et al., 1982, 1986) and the other source to have provided a useful metaphor to think about the idea of proceduralism (Crowder, 1989). Together, they provided me with guides to move to an emerging construct of implicit social cognition.

A first foothold along slippery terrain was Russell Fazio’s work on automatic attitudes. Almost 2 decades ago, Fazio et al. (1982) made a unique connection between attitude and memory through the construct of accessibility:

Consequently, the present paper focuses upon a somewhat different indication of the “strength of an attitude”—the speed and ease with which the attitude can be accessed from memory. Attitude accessibility is a particularly relevant and important variable, for, unlike other potential indicants of attitude strength, attitude accessibility is obviously involved in the attitude-to-behavior process. If an attitude is to “guide” behavior, the attitude must first be accessed from memory. (p. 340)

Later in the same article, they stated for the first time to my knowledge what is now the received view of automatic attitudes (at least among those who believe that evidence for them exceeds that for cold fusion): “Essentially, then, an attitude can be viewed as an association between a given object and a given evaluative category” (p. 341).

I regard this statement as marking a leap forward for the study of attitudes and for the study of all related mental constructs. Why is this? Because unlike previous definitions that had perhaps assumed but never directly stated it, Fazio et al. (1982) took a measure of the strength of evaluative associations to be a measure of attitude. Later, Fazio et al. (1986) accomplished a shift in the paradigm through a series of demonstrations of the automatic association between attitude object and evaluation. In so doing, they made further debate and development possible (Bargh et al., 1992; Glaser & Banaji, 1999). Fazio’s work allowed me to shed my orthodox assumption that attitudes were measurable only in conscious form and, more important, it provided a way to conceive of evaluative associations as an individual difference measure of attitude valence and attitude strength.

A second source of influence that dragged me out of my orthodoxy was a simple analogy offered by Bob Crowder in his attempt to capture the idea of proceduralism. In an essay honoring the contributions of Endel Tulving (Crowder, 1989) he wrote that

once memory is regarded as a by-product of information processing, the implicit concept of an all-purpose memory store dissolves. To me, a comfortable language to describe this attitude is to say that memory is not a storage process as such; it is simply the property of information processing that extends in time afterwards. In much the same way, the neutrinos detected at several international observatories starting on February 24, 1987, were not a sort of time capsule, laid down by Supernova 1987A for our benefit. The neutrinos represent part of that original event itself, in a galaxy called the Large Magellanic Cloud—an event that occurred 163,000 years ago and a billion billion miles away from here. And so it is with at least some kinds of memories. The retention is just an aspect of the original episode itself, manifest at some temporal remove. (p. 272)

This way of thinking about memory, not as a storage device but rather as a property of information processing, detected at a time and in a place distant from the original episode (in response to an appropriate probe) provided me with the brace to view apparently different constructs in a similar way—attitudes, beliefs,

and their by-products in social cognition. It seemed that just as it had become possible to think about memory not as a store in which memories lay waiting to be selectively (re)collected by the mind's conscious hand, so it should be with at least some kinds of attitudes and beliefs. A few years later, we wrote about implicit attitudes in the same spirit of proceduralism (Greenwald & Banaji, 1995), and it has been central to our work on implicit social cognition with special focus on prejudice and stereotypes. Not surprisingly, because the terms *prejudice* and *stereotype* have clear meaning in ordinary parlance, they resist molding (even more so than the parent concepts of *attitude* and *belief*) in all ways that violate the texture of their everyday, colloquial meaning. But surely this should be regarded as an impediment, and as such, it requires correction, if these constructs are to be treated not on the basis of how hard they tug at one's intuition but as the objects of scientific scrutiny.

Déjà Vu All Over Again

The leap forward that is required in understanding implicit attitudes is no different from the advances in measurement in the early decades of the 20th century. In 1928, the idea that attitudes could be measured was novel enough that Thurstone (1928) needed to make an argument for it by drawing a parallel to the measurement of the dimensions of a physical object, a table. The title and tone of his article show clearly that his ideas concerning measurement were not commonly accepted. In particular, a footnote shows that there may have been resistance to his proposal that what he was measuring was indeed an attitude. Here is a part of the footnote: "Professor Faris, who has been kind enough to give considerable constructive criticism to the manuscript for this paper, has suggested that we may be measuring opinion but that we are certainly not measuring attitude" (Thurstone, 1928, p. 532).

Thurstone (1928) then went on to argue in the footnote why it is the case that what he was measuring was not an opinion but an attitude. A century ago the very idea of measuring evaluation or attitude was sufficiently radical to provoke disagreement about whether this new mental contraption was indeed reflecting an "attitude." I should therefore suppress my surprise if the idea of unconscious attitude encounters similar doubt, this time because speeded judgments that tap the strength of evaluative associations simply do not intuitively capture attitude as we have come to (not) know and love that construct. The process, the contemporary Professor Faris said, may be measuring something, but it is certainly not measuring attitude: "More generally, there is little evidence to suggest that the IAT has anything at all to do with attitudes, either implicitly or explicitly" (see Karpinski & Hilton, 1999).

There may be two lessons from this history for students of implicit attitudes and implicit social cognition more generally: First, in addition to cogent theory, the study of implicit attitudes needs to be deeply grounded in an understanding of the psychometric properties of new measures and in setting rigorous standards for

measurement in a manner similar to the first century of research on explicit attitudes.² Yet at the same time, understanding the nature of these new measures cannot come from an unthinking adoption of criteria developed for the different family of explicit measures. Second, it is necessary to resist any and all orthodoxy that challenges developments of the concept of implicit attitudes on the grounds that the measures do not bear a likeness to instinctual accounts of how attitudes should be measured or how they have been measured in the past. To put it baldly, that sentiment is as misplaced today as it was in 1928.

It is perhaps the case that early in the development of any construct, there are moments of understandable and even required confusion about the appropriate manner in which the construct should be treated. With psychological constructs (vs. many physical ones), the problem has an added peculiarity identified by William James in his discussion of the scientific study of self, that is, that the knower and known are one and the same. For many constructs like memory, attitude, or consciousness, the knower and known have a close relationship, a condition ripe for delusions that derive from intuitive notions of what memory, attitude, or consciousness ought to be. Among the most invidious of traps when investigating mental constructs with which one's own thinking apparatus has intimate familiarity is the demand that the constructs ought to feel "real," even to the scientist. Several years ago, Bob Crowder and I noted what appeared to us to be such a peculiarity in the study of memory: the demand that the workings of memory ought to be studied in their natural habitat because of an assumption that methods that were ecologically valid would naturally yield results that are more valid (Banaji & Crowder, 1989). To the contrary, our view was that the appearance of the ecological validity of the probe used to study memory ought not to primarily guide investigations, in spite of the comfort that accompanies such a strategy.

A similar argument, quoted here in full, was made by Crowder in a commentary titled "The Brain, the Kidney, and Consciousness" in the early 1990s of which I retain only a draft copy that was circulated for comment by him.³ (Crowder's mention of consciousness here refers to intuitive theories of how mental systems work; he is not against the scientific study of consciousness.)

²That is why Greenwald et al. (1998) published the first article years after the task was first designed. Hundreds of people have now participated in dozens of parameter-testing studies at the University of Washington and Yale University, conducted not for publication but for understanding. I warn against the use of implicit measures without a deep understanding of them (Karpinski & Hilton, 1999, being the best example); the technique can appear to be easy to construct.

³As Bob Crowder's students recently discovered, he did not always enter citations to his published work in his curriculum vita (a fact I attribute to his intense modesty rather than disorganization). I have been unable to locate the full citation to this commentary, although I am certain that it was published. I am grateful to Julie Crowder for locating the most recent version on Bob's computer—the citation here is from that version.

Consciousness of the mind is precisely the same as consciousness of kidney function. Both the brain and the kidneys are organs that result in bodily processes. Leaving aside truly theological issues, they could not, possibly, be anything else, other than bodily functions. For both, we can observe the states that result from activity in the system, but we have no special access to how either works. We have only the evident products of each to consider and from which to fabricate models of process. Naïve theoretical models result from observations of the products of the system and scientific models (right or wrong) result from scientific methods.

Our intuitions based on experience, or consciousness, are of equal value for the study of mental and renal function. They are not, by the way, of no value at all: We laymen carry with us a naïve theory of kidney function based on observed input-output relationships. I call this the "asparagus and beer" process. It can lead to some valid conclusions, with careful enough observation and enough time. But it is entirely extracurricular with regard to the agenda of natural science. The development of kidney physiology depended surely on histology, anatomy, and physiological laboratory experiments, not on more and more thoughtful theories based on kidney consciousness.

Experimental psychology seems to be the appropriate natural science to investigate how the mind works. The study of consciousness of mental life will inform ideas on how the brain works no more than awareness of other bodily functions informed their corresponding branches of science. To deny this means to accept that the brain is translucent in process whereas other organs are opaque. Why should this be the case?

Are kidneys too banal a system to use as an example? Then take vision, another bodily function resulting from identified physiological substrates. Indeed, this is an excellent example because retinal tissue is authentically a part of mind. The careful analysis of visual consciousness—after images, contrasts, illusions—has occasionally informed visual science, but experiments are usually needed to bring these phenomena under control, for they are, after all, largely unconscious.

In auditory cognition, we simply cannot become aware, however hard we try, of the cues that guide perception of things like speech (formants). Nor can we count the vibrations that make up periodic sound. This said, can we be optimistic about the contribution auditory consciousness could make to our understanding of audition? Would we want to deny that auditory cognition is mental? These systems, kidney function, visual function, and auditory function, can serve as parables for how science can deal with mundane experience: Above all, by abandoning that experience and using the methods of natural science to find more adequate ways of describing how the system works. This recommendation is no different in the mental domain than in the renal. Consciousness of any of these systems comes "after the fact" and cannot clarify the process.

In one regard, Crowder and I were "kindred spirits," an honor he once heaped on me that marked the beginning of humorous written exchanges that resulted in several published works (see Banaji & Crowder, 1989, 1991, 1994). We were both prone to speaking our mind about a need for the ordinary, and hence respectful,

treatment of psychological constructs, and we were prone to stating such opinion, to put it politely, in plain language. In this next section, I comment on work that, in one sense, has nothing to do with the work of my kindred spirit and, in another sense, has everything to do with him because the approach and interpretations have required speaking plainly, and the encouragement to do so was inspired by Crowder's own unvarnished words about matters such as kidney consciousness.

Are Implicit Attitudes Valid?

Prediction, Prediction. Did You Hear Me? Prediction!

Some constructs elicit curious early demands of a particular form of validity, especially criterion validity, both of the predictive and concurrent sort. This demand is, for example, more true of constructs like attitude (not to mention prejudice) rather than, say, force (not to mention tension, surface tension, or hypertension). The insistence with which the "validity" question regarding a concept and its measurement is raised may be inversely related to the "toughness" of the concept's scientific standing—the stronger the assumption that a concept has already earned admission to the mansion of science, the lower may be the demand to immediately prove just what the particular procedure and the construct it presumably captures really predict.

The problem of validity shows up in at least two ways. First, there is a demand to show whether the measure predicts some criterion. For example, if the measure were one that measured romantic feelings, the question that sometimes arises quickly is how well it can predict Joe Schmoe's problem of a broken heart. In fact, the criterion variable of fixing the broken heart assumes great importance, often without regard to the question of how well understood the criterion (heartstrings, heartthrob) itself is. This is an important issue because as McGuire (1989) pointed out, we as a field have a far better grasp over matters of mind (i.e., mental constructs like attitude or memory) and their measurement than we do over behavior and its measurement. In addition, the fact that the criterion (often a behavior that is of the "real world") is often far less understood can make measurement of the criterion itself, let alone the relationship, weak. Relationships in areas that have an underdeveloped understanding of the behavior have a low probability of being detected (even if they exist), and quickly, in consequence, a sense can develop that the construct is not valid because of poor prediction. When done right, mental states will predict behavior (see Bargh, Chen, & Burrows, 1996), but to understand the mental construct, the behavior, and their relationship is not a simple or quick task.

We as a field did not ask in the early 1970s what lexical decision tasks predicted about other forms of automatic memory, and similarly, we did not ask in the 1980s what word-stem completion would predict about anything at all. Certainly, there has been no scurrying around to test whether performance on a word-stem completion in

the "lab" would predict how well Jane would do on her fifth-grade fill-in-the-blank test in the "school." That attempt would be preposterous because the purpose of a task such as stem completion was to understand something about how memory works. It is equally premature to clamor for immediate criterion validity of evaluative priming, the evaluative Stroop, or the IAT at this stage. In other parts of social cognition, I have no recollection of similar demands being made. For example, in the past two decades of important research programs provoked by Higgins, Rholes, and Jones (1977) on construct accessibility, there was no great criterion validity crisis of whether the "Donald" task (a dependent variable that assessed the degree to which a primed construct influenced judgment) predicted impression formation at cocktail parties. Pushing the criterion validity problem early and strongly can have at least two negative outcomes: It can stop research through early dismissal of a tool, and it can provoke misuse of the construct based on early and insufficient tests of predictive validity (a case in point being the now multimillion dollar international enterprise of emotional intelligence testing based on a shamefully nonexistent understanding of the construct), a point the psychologists who developed the original construct have repeatedly emphasized (Salovey, Mayer, & Caruso, *in press*).

The task of building the case for a phenomenon using a particular task and through comparisons with other procedures is a slow and hard process. Fazio (*in press*) offered a brief review of largely theoretical issues involved in investigations of evaluative priming in which an underlying theme becomes clear—there is a seamless connection between explorations of phenomena that have a decidedly theoretical focus (what are the competing theories, mechanisms, moderators, etc.)—and many demonstrations show the validity of evaluative priming in an incremental way—some show how basic cognitive processes such as attention, categorization, judgment, and behavior are affected by automatic evaluation. Some studies clarify the construct validity of automatic attitudes using the evaluative priming procedure. Later, studies show concurrent criterion validity by emphasizing that the strength of the automatic attitude activation effect using racial stimuli predicts the magnitude of nonverbal prejudicial behavior. But understanding automatic attitude or any such fundamental process takes time. For example, the idea of an automatic attitude-evaluation link was first described in 1982 (Fazio et al.), whereas the famous criterion validity study was conducted over a decade later (Fazio et al., 1995).

Over the course of that decade, a great deal more of the "nomological net" described by Cronbach and Meehl (1955) to track the development of the validity of a construct was carefully constructed for evaluative priming as a measure of automatic attitude. The tolerance sought in the assessment of a new measure of the same underlying construct, the IAT, is the same—the time to develop the meaning of what a task, with its unique properties, can teach about unconscious attitudes. Pushing fast and furiously to "show me what it predicts" may be counterproductive. One first needs to understand the construct before asking what it may or may not predict. The fields of intelligence testing and of testing in selection contexts such as

organizations highlight some of the issues here, and profound discussions concerning validity may be found in Jenkins's (1946) "Validity for What?" in Weitz's (1961) "Criteria for Criteria," and in Wallace's (1965) "Criteria for What?" These may be worth paying attention to before issuing subpoenas.⁴

But why is this demand for prediction in some domains more so than in others? One obvious possibility is the presumed social value and social interest of the particular construct under scrutiny. If social interest in the topic is high (e.g., stereotypes, prejudice), questions concerning prediction arise faster than one can say "Jackie Robinson." As hard a realization as this may be for those who study memory, people do not give as much of a hoot about implicit memory as they do about implicit prejudice. This was perhaps the fortunate circumstance in which even the construct of automatic attitudes, when studied outside the race-prejudice context, was able to prosper. Attitudes about neckties grab about the same attention as does memory for neckties. So the difference here may be less in the construct (attitude vs. memory) than in the content of what is being measured—good reason for closely monitoring the sources of displeasure or impatience. Yet unless attitudes that reveal social group preferences are subjected to questions of criterion validity in a manner no more or less urgent than the same questions about the predictive validity of blood flow as a measure of brain activity, or response time as a measure of memory, one endures the risk of driving the best minds away from fields where exploration of socially relevant constructs needs to become accepted as an intrinsic, scientific good.

In that same 1928 article with which I began this chapter, Thurstone said the following about his interest in attitude measurement, even if no prediction of behavior were possible:

In the present study we shall measure the subject's attitude as expressed by the acceptance or rejection of opinions. But we shall not thereby imply that he will necessarily act in accordance with the opinions that he has endorsed. Let this limitation be clear. The measurement of attitudes expressed by a man's opinions does not necessarily mean the prediction of what he will do. If his expressed opinions and his actions are inconsistent, that does not concern us now, because we are not setting out to predict overt conduct. We shall assume that it is of interest to know what people say that they believe even if their conduct turns out to be inconsistent with their professed opinions. Even if they are intentionally distorting their attitudes, we are measuring at least the attitude which they are trying to make people believe that they have. (p. 533)

⁴My collaborator, Tony Greenwald, while agreeing with my sentiment, thinks differently. He believes that reasonable (and unreasonable) demands for predictive validity can have positive consequences for the field in the form of initiating and completing research faster than in the absence of such demands. He is a better person than I.

Dreading the Nonintuitive

Another sense in which the validity problem arises is to assume that the measure should "feel" valid. A statement such as "I used to be an atheist, but I gave it up for Lent" (attributed to the character Max Klinger of the television series *M*A*S*H*) intuitively feels like a statement of a personal belief (and belief change) in the same way that asking for free recall intuitively feels like a measure of memory. To the contrary, a task that measures the strength of association in response latencies between me + atheist has less of that intuitive advantage as a measure of implicit belief, in the same way that measuring the strength of association between doctor + nurse seems intuitively less to be a measure of semantic memory. Justifying a method of exploration on grounds of such intuition was exactly what Crowder was complaining about in his piece on kidney consciousness.

Yet I hope it is clear that I am not saying that the constructs one should engage cannot be ones that already have lay meaning. What I am saying, especially when it comes to constructs like attitude and prejudice, is that one proceeds with validation without a simple judgment of the similarity between the surface features of the measure and its likeness to the intuitive representation of the construct. It is "work" to write papers, but that lay sense of what "work" means is neither expected to, nor does it, bear isomorphic mapping with the same construct as defined in the science that has studied it. As millions of students have memorized for physics exams, work is said to be done when the point of application of a force moves (Sears, 1980; but see Feynman, 1963, for a discussion of the difference between physical work, physiological work, and the phrase "workers of the world, unite!," pp. 14-1-14-2). I think people have inadequately come to terms with the idea that constructs of attitude and prejudice are objects of scientific scrutiny. If they had, they would make less of a demand that a measure of attitude must feel like a measure of attitude, by first passing some personally and communally comforting threshold of validity.

It may be best to turn to what is already known about the particular task in question (the IAT) and ask whether in the few years it has received attention, progress has been made to form a strong first stage of the "nomological net" set forth as the requirement for developing construct validity (Cronbach & Meehl, 1955). To address questions about the psychometric properties of data from implicit measures, I begin with a description of the findings themselves. In the course of doing so, responses to the question of validity are intermixed and evidence is provided that rules out some alternative explanations, leaving others as contenders for further experimentation. The question of validity is a slippery one, in part because there is no clear adjudication of a measure as valid or invalid, just those that have garnered greater or lesser evidence in support. Cronbach and Meehl (1955) suggested that often an investigator believes that no single criterion is available to meet his or her standards of perfect criterion validity and that, under these circumstances, an interest in construct validity must come to the fore. It is the trait or quality that

underlies the measure that is of central importance rather than the test behavior or the score on the criterion variable.

In a previous section, it was noted that in studies of semantic priming, while speaking about strength of association as the immediate explanatory device, investigators have been clear in their assessment that their data spoke to the question of semantic memory. Cronbach and Meehl (1955) offered a direct general statement that addressed this issue: The question of construct validity would not arise if the goal was merely to say that the pattern of response latencies on a lexical decision task reveals faster responding to concepts that are semantically related, or that the evaluative priming task captures faster responding to tasks that are affectively related. No question of construct validity would arise because "no interpretation has been made" (p. 283). However, as soon as such data are interpreted to reflect memory or to reflect attitude, the question of construct validity becomes important.

Furthermore, if one were interested in speaking to a single practical question, a criterion variable may be easy to set up and appropriate tests of prediction may be conducted. But if one labors under the illusion of building general theories of social cognition (as do many psychologists, including my colleagues and I), one must remain engaged with the question of construct validity no matter what the demands are to do otherwise. Drawing from Thurstone, Cronbach and Meehl (1955) wrote,

it used to be common to define validity as the correlation between a test score and some outside criterion. We have reached a stage of sophistication where the test criterion correlation is too coarse. It is obsolete. If we attempted to ascertain the validity of a test for the second space-factor, for example, we would have to get judges [to] make reliable judgments about people as to this factor. Ordinarily their [the available "judges"] ratings would be of no value as a criterion. Consequently, validity studies in the cognitive functions now depend on criteria of internal consistency. (p. 283)

Lessons learned 40 years ago seem to be lost in the demand for a single type of (predictive) validity.

Toward Validation of the Implicit Association Test: Implicit Race Attitudes

This last section presents a thin slice of a developing net of results, many of which are only recently available (for a more general review of the literature on unconscious social cognition, see Banaji, Lemm, & Carpenter, in press). The findings presented here address the challenges that have been posed to the IAT, at least with a single attitude object, race, with validity in a variety of ways already available. Early in its development, the IAT technique seemed both to be of scientific interest while also providing a uniquely captivating experience. It was capable, at times, of dramatically

highlighting the ease of some associations over others, much in the same way that the Stroop task leaves the respondent cognizant of the response competition that hinders some associations compared with others. For these reasons, the developers of the IAT decided to make the task available in more than one way: first to other scientists for research use in 1995, even prior to publication of the first article (Greenwald et al., 1998). Second, after approximately 4 years of experience and dozens of parameter-testing experiments in two labs (Yale University and University of Washington), a website was created at which various tasks can be sampled primarily as an (un)consciousness-raising device⁵ for the public at large (<http://www.yale.edu/implicit>).⁶ By April 2000, well over 700,000 IATs using a variety of attitude objects were completed (see Nosek, Banaji, & Greenwald, 2000, for a summary).

In my mind, the appeal of and perhaps resistance to the IAT derive in part from the relatively stark experience the task provides to the respondent of the mental interference that makes some associations (e.g., Black + good) harder to perform than others (White + good). The discomfort in the case of the race task may stem from exactly such awareness of (a) an evaluative disparity between conscious attitudes and unconscious evaluations and (b) a lack of control over one's responses on a task that has personal meaning and value. Certainly the developers of the technique were themselves taken aback by their own first performance because their data on this measure sat in direct contradiction to their presumed attitude. From the various responses to date, it appears that acknowledging that both the consciously available and the unconsciously elicited attitude can be true renditions, each of a different form of the same attitude object and within the same mind, appears to be an acquired taste.⁷

When data from amnesia patients or college students show stark dissociations between conscious and unconscious memory, one can see this as theoretically

⁵John Jost first proposed the use of the phrase "(un)consciousness raising" (Jost, Banaji, & Greenwald, 1994).

⁶The creation of this website would not have been possible without the expertise and determination of Brian Nosek, currently a graduate student at Yale, who serves not only as web master and primary data analyst but also as the architect of more general possibilities for web-based research.

⁷Roddy Roediger pointed out that implicit attitude research appears to have constructed a debate about whether such attitudes reflect "real" attitudes more so than do explicit ones. To him (a favorer of transfer appropriate processing), such an issue seems to be a nonstarter. Both measures can be valid indicators of different types of attitudes and behavior (personal communication, January 25, 2000). I could not agree more. In fact, when considering how individuals or societies must prepare themselves for decisions of great personal or societal consequence or in building aspirations for the future, the more I learn about the presence of implicit social cognition, the more I am struck by the importance of conscious thought and action in human affairs (an idea to be developed more seriously in a DeVane Lecture, the collaborative course being devoted to the future of American democracy, to mark Yale's tercentennial in 2001).

exciting and as marking a major step forward in understanding. When the vast majority of respondents who hold consciously egalitarian attitudes (the theoretical equivalent of amnesia patients for the purpose because they can reveal disparities between conscious and unconscious attitude) show marked dissociation between conscious and unconscious attitudes toward Black and White Americans (and in several other attitude conditions), the occasion should be cause for similar theoretical interest and debate. I need to say something banal here, in the sense that most all agree with it: Tasks and findings ought not to be judged by the level of comfort they provide about acceptable views of human nature or the personal comfort the findings afford. John Bargh (1999) made the wise observation that just because we as humans may like the idea of being able to exert control over stereotypes and prejudice is no reason to believe that we can or to reject findings that show that we cannot. It is not surprising when a lack of synchrony between our view of ourselves as unbiased ("I am a morally good person") and evidence of ourselves as biased ("I am not a morally good person") controls assessments of the finding of such discrepancies on the part of laypeople; it is a bit embarrassing when the same is possibly true of scientists themselves.

This chapter contains descriptions of laboratory data in the usual sense but also e-data obtained at a website that opened in September 1998 and at which various demonstrations that tap implicit attitudes and beliefs can be sampled (for details, see Nosek, 1999; and Nosek et al., 2000). These particular e-data are interpreted with caution because participants were self-selected visitors to the website, who self-selected the tasks to complete and the number of times they completed each task. However, these e-data have a unique strength because of the sample size of completed tests (well over 1,000,000 tests completed between September 1998 and December 2000). Additionally, the e-data provide a sample that is in some regard more diverse than the college samples typically tested (see Nosek et al., 2000). Reassuringly, the patterns of e-data collected at the demonstration site conform excellently to data collected under more controlled conditions, and they also conform to theory.

Automatic Preference for White

There are several findings of interest in the race data, shedding light on the role of group membership and the cultural value assigned to social groups in determining attitude. Several independent data collections show a strong pro-White preference⁸ among White Americans, even when using conservative cutoffs for computing prefer-

⁸For ease of description I refer to findings as revealing attitudes that are pro and anti toward the attitude object. In all contexts, a proattitude object bias (e.g., young) refers to a greater ease of pairing that category (young) with evaluatively positive concepts (good or pleasant words or pictures) than negative concepts (bad or unpleasant words or pictures). All reports here are based on a task that always includes a contrast category (e.g., the elderly population), as in the original specification of the IAT. An alternative task that avoids the inclusion of a single contrasting category is currently being developed.

ence. The original article by Greenwald et al. (1998) first shows this pattern, and it has now been obtained in every data set they have collected since then, with effect sizes (Cohen's d) that are usually in the large range (.8 and greater).

Beyond Familiarity

The most common early alternative posed in response to these data from White (and other non-Black) respondents concerns differential familiarity. The pro-White preference, it can be argued, really reveals an effect of relatively greater familiarity with White than Black stimuli, rather than reflecting anything like an implicit preference for one group or another. This issue has been addressed by Dasgupta, McGhee, Greenwald, and Banaji (2000) and Ottaway, Hayden, and Oakes (in press). Even when Black and White names are statistically equated for familiarity, a preference for White over Black remains, suggesting that familiarity cannot explain the obtained pro-White preference. Additionally, using multiple social groups, Rudman, Greenwald, Mellott, and McGhee (1999) showed that the IAT effect is not an artifact of familiarity.

Culture or Person?

The finding of a pro-White effect among White Americans has persistently raised the possibility that what the IAT detects is not a reflection of the individual's own implicit attitude but rather a preference that resides in some clearly separable culture out there. If culture is offered as the source of learning that determines the content of the automatic preference, there is only agreement. But if the measure is assumed to have detected something other than that individual's temporary representation of the attitude object (i.e., it is seen as a measure of something else called "culture"), I would disagree. There is no question that the IAT (like other attitude measures) reflects a learned preference as dictated by one's culture and group membership. For example, a semantic priming task roughly detects repeated cultural pairing (moderated through individual experience) of, for example, the concepts *doctor* and *nurse*.

In the same way, the IAT roughly detects repeated cultural pairing (moderated through individual experience) of Black + bad and White + good, most clearly among non-Black inhabitants of the United States. But just as the strength of association between doctor and nurse in a given person reflects how those constructs have come to be paired in the mind of a particular individual, so does the variation in pro-White bias reflect the strength of association between White + good in an individual's mind, however culturally "caused."

The following example should clarify the reason for the mistaken belief that the preference being measured has little to do with an individual's preference. It should surprise no one when I say that it is through cultural learning that children in South India learn to eat and love very hot pickles (even though all infants,

including South Indian ones, spit them out with vigor). What is interesting is "whose" attitude toward pickles one then believes the eventual adult attitude to be. I would argue that one sees this attitude as belonging to the individual (i.e., as Suparna's attitude or Kavitha's attitude), however obvious may be the cultural influence. As a field, we believe that attitudes, although showing cultural variation (e.g., some Americans liken the taste of Indian pickles to that of gasoline, whereas millions of Indians cannot get through a meal without them), also reflect the attitudes of individuals embedded in that culture. In addition, to social psychologists, it is the individual differences in those attitudes that are important and interesting, in addition to group differences. Indeed, it is individual variability that is at the core of the construct of attitude.

This being the case, and an agreeable issue when one speaks about pickles, it is interesting that the IAT-detected attitude is said not to reflect an individual's implicit attitude but rather a cultural bias that is independent of the individual's own attitudes. For example, "in other words, perhaps the IAT tells us little about people's conscious or unconscious beliefs or attitudes, but tells volumes about their environments" (Karpinski & Hilton, 1999). To my mind, the idea that the IAT (or evaluative priming) attitudes are seen not as a reflection of an individual's mind is at best an oddity that will prove in time to be misguided.

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

But the more important observation here may be the following: The experience that implicit attitudes, as measured by the IAT, may not reflect an individual's own attitude but rather that of the culture may lie in the dominant popular understanding of attitudes—as things that are under conscious awareness, intention, and control. In addition, this is a meaningful experience and a distinction that consciously held attitudes certainly allow. That is, one can consciously have the compelling experience of holding a belief or attitude that is discrepant with those of individual others (e.g., "My senator likes the NRA, but I don't") or beliefs that are discrepant from a culture or subculture (e.g., "97% of all Americans and 95% of physicians believe in God, but I don't"). The human ability to consciously "know" one's own attitude or belief, and to "know" its separation from the attitudes and beliefs of others, is an important marker of conscious social cognition. The ability to be able to consciously reflect on one's own mind, a fundamentally unique human ability, is what appears to be causing the confusion regarding implicit attitudes. One desires to see a separation between culture and person in the same way with implicit attitudes as with explicit

attitudes; this distinction is imposed on the data, so powerful is the assumption of individual–culture separation (for a clear example of this fallacy, see Karpinski & Hilton, 1999). The expectation is that just as conscious attitudes are malleable by volition, so must be the case with automatic attitudes. When implicit attitudes do not respond to the call of free will, the source of the attitude becomes suspect—whose attitude is it? “Not mine,” is the answer. “I can’t seem to control it, and surely if it were mine, I would be able to do so.” Add to this the unpalatable nature of the observed dissociation between conscious and unconscious race attitudes, and one may see why a manufactured distinction between self and culture can seem so compelling, even if incorrect.

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

Is Group Membership All That Matters?

Among the most intriguing of questions concerning implicit race attitudes concerns the performance of Black Americans themselves. The discussion rests primarily on e-data because of the sheer size of the numbers of Black Americans available at the demonstration website. Black Americans show a notable deviation in automatic attitude from White Americans, with pro-Black preference being revealed to a greater extent than it is among White Americans. However, Black American data do not show strong pro-Black preference. Black Americans show an average difference score close to zero (no bias in either direction). These data support a theoretical expectation I had about Black Americans. To the extent that the IAT captures a preference that stems from group membership (rooted in a preference for self and hence attributes associated to the self, see Greenwald et al., 2000), Black Americans show a more positive automatic attitude toward the group to which they belong than do White Americans, who do not possess group membership. Yet to the extent that the IAT reveals a learned negativity toward Black Americans that is inherent in the culture, approximately half the members of the group of reported Black Americans in this sample do not (or one might say, cannot) show implicit positive attitude toward their own group as do members of many other more culturally privileged groups. This is in stark contrast to the strong progroup attitudes that Black Americans show on explicit measures that are even stronger than those shown by White Americans (see Banaji, Greenwald, & Rosier, 1997). So group membership does matter, but

because all groups do not show equal implicit positive associations to their own groups, I entertain the possibility that culturally imposed notions of good and bad infiltrate and shape individual attitudes toward one’s own social group.

Third Parties

The influence of culture on individual attitudes may be gleaned in another way—from third-party members. If attitudes are a function of both group membership (i.e., preference for one’s own group) and cultural value assigned to the group, then respondents who belong to a third group (i.e., those who do not hold membership in either of two groups) can provide useful information about the reach of culture. Asian participants provide useful data in this regard. Of the 8,012 Asians who took the Black–White IAT, 73% showed a pro-White bias, that is, an effect commensurate with that shown by White Americans. Likewise, with the many implicit attitudes studied that cannot include group membership (e.g., insect–flower, math–arts), the attitude may be informative about individual differences toward the object, even though no membership is implied. At times, even here, known-groups validation is available, as in the finding that entomologists show greater liking for insects than do nonentomologists (Citrin & Greenwald, 1998), and omnivores show greater liking for meat than do vegetarians (Swanson, Rudman, & Greenwald, in press). More than anything, this finding does verify that the long reach of culture affects even third parties who do not hold membership in the target groups.

Social Neuroscience: Attitudes and Brain Activation

In a previous context, I mentioned Robert Crowder’s belief that the brain, like other organs involved in bodily functions, should be studied through experimental treatment of behavior and not from deeper and deeper intuitions about the working of the organ. In the context of research on attitudes, Phelps and I postulated that dissociations between implicit and explicit attitudes observed on measures of automatic associations should find validation in brain activity (see Ochsner & Lieberman, 2000, for a review of extant research on social neuroscience). In particular, there ought to be a mapping between individual learning (as revealed by IAT and other measures) and activity in a subcortical structure known to be involved in emotional learning. Phelps et al. (2000) focused on the amygdala, obtaining a measure of activation in response to unfamiliar Black and White faces during scanning. The difference in Black–White activation was significantly correlated with measures of race bias obtained on the IAT ($r = .58$) and potentiated eye-blink startle ($r = .56$) that also used unfamiliar Black versus White faces. Two findings of a lack of correlation are also of theoretical interest. First, such correlations were not obtained when faces of familiar and liked Black and White individuals were

used. Second, no correlation between amygdala activity and a measure of conscious race attitudes and beliefs, the Modern Racism Scale (MRS), was obtained. Their point is the opposite of the simplistic notion that "racism is in the brain." Instead, they suggested that cultural learning is the cause of the effects observed in both amygdala activity as well as IAT and potentiated startle. For the purpose of this chapter, the finding of interest lies in the congruence between the responses of a brain center known to be involved in emotional learning and the automatic expressions of preference on "behavioral" tasks such as the IAT and startle. Culture is the ether in which learning occurs, and such learning reveals something specific about the internalized attitude of the individual; searching for a ghost named "culture" in the mind's machine will likely produce the same successes as other searches for ghosts.

Ethnocentrism

With its roots in the postwar exodus to the United States, scholars especially from the Frankfurt School favored a theory regarding a complex of mental and social processes that produced a personality type they referred to as "the authoritarian personality," a signature feature of such a personality being ethnocentrism, or a generalized dislike of all groups foreign to one's own. The notion of authoritarian ethnocentrism produced a great deal of research, but such work did not touch the mainstream of social psychology nor did it leave a lasting impact on the field. Cunningham, Nezlek, and Banaji (2000) analyzed ethnocentrism using both classic explicit and IAT measures, including five social groups for assuring generality of the findings—attitudes toward Black–White, gay–straight, poor–rich, foreign–American, and Jewish–Christian. This research shows that the strength of automatic evaluations of good–bad with each of these pairs reveals not only a stable implicit ethnocentrism factor but also one that is psychometrically distinct from the latent factor that taps explicit attitudes. This work marks the beginning of a program of research to provide construct validity for implicit measures with both convergent and divergent measures. By including measures of general rigidity in thinking and authoritarianism this research shows that unconscious attitudes and broader personality structures operated in concert.

Psychometrics of Implicit Attitude Measures

Many conversations at conferences in recent months have focused on the problem of the lack of relationship among various measures of implicit attitudes and between measures of implicit and explicit attitudes (e.g., National Science Foundation Construct Validity/Implicit Social Cognition Workshop, May 1999, Zion, IL). These questions lie at the heart of many of the challenges that are being posed to implicit measures of social cognition more generally. Some investigators have selected a single explicit measure of attitude as the criterion (this itself may indicate the

dominance of explicit measures, that they are naturally assumed to be the criterion!) and attempted to see if the implicit measure "measures up" to it. For example, Karpinski and Hilton (1999) did just that, arguing that if the attitude topic is one not affected by social demands, implicit measures ought to relate to explicit ones. That such an assessment was offered only indicates my own ineffectiveness in pointing out the expected dissociation between such measures, even when social desirability is not at issue. One can see these two families of measures (implicit and explicit) as distinct not merely as an empirical convenience (i.e., experimenters need to keep the hypothesis hidden from participants' awareness, so they must use measures that bypass such awareness) but as theoretical necessities; that is, these two modes are fundamentally different, even when they are partially related (Cunningham et al., 2000; Greenwald & Banaji, 1995).

The facts as they currently stand suggest that implicit and explicit attitudes are indeed dissociated and sometimes doubly (e.g., positive attitudes on an explicit measure, negative attitudes on an implicit measure) or they may be partially dissociated (both attitudes may be positive or negative, yet each accounting for a unique portion of the variance). In the research world in which conscious and unconscious attitudes are seen to be conceptually distinct, the low correlation or lower correlation between measures within the same family is taken as evidence of validity, not a challenge to it. Thus, when Fazio et al. (1995) found a lack of correlation between evaluative priming and the MRS, a theoretical prediction was confirmed; when Phelps et al. (2000) found that IAT and brain activity were correlated but that MRS and brain activity were not, convergent and incremental validity were demonstrated; when Cunningham et al. (2000) showed that implicit attitudes toward five social groups formed a single factor that was separable from (but related to) explicit attitudes, convergent and divergent validity were demonstrated.

Just as across family (implicit–explicit) comparisons are of interest, so are within family (implicit–implicit) comparisons. A worry is that various versions of implicit measures, theoretically belonging to the same family, are not related to each other. Certainly, if that were the case, the outcome would be similar to research on memory: Multiple implicit measures of memory tapping differing processes are not expected to be related to each other even though they sit theoretically in roughly the same family. With attitudes, such data have begun to be reported (e.g., Cameron, Alvarez, & Bargh, 2000). Nevertheless, it is likely that low and unsystematic correlations are more likely to be a function of the lack of power of experimental designs and statistical analyses rather than reflecting a genuine lack of correlation. It is reckless to affirm low or no correlations without effectively ruling out Type II errors (see Cunningham, Preacher, & Banaji, in press). Reaction time data are notorious for insensitivity to internal consistency, but when internal consistency is controlled for, the stability of the data over time increases dramatically. Various researchers have conducted studies in which multiple measures (across time) of multiple tasks (all variations of implicit attitude measures) are used and good convergence between

implicit measures is found (Cunningham et al., in press). Why is this? Unlike other tests, the substantial power both in sample size and data analytic techniques has been ensured, calling into doubt one's own and others' reports of a lack of convergence within implicit measures (see also Rudman & Kilianski, 2000). The moral of the story here is that it is too early to proclaim dissatisfaction with the reliability or stability of implicit measures. The overwhelming likelihood is that, when appropriately conducted, theoretically expected relationships between implicit measures indeed emerge.

Shifting Implicit Attitudes Follow Shifting Construals

Mitchell, Nosek, and Banaji (1998, 2000) used explicitly liked Black athletes (e.g., Michael Jordan) and explicitly disliked White politicians (e.g., Jesse Helms) as stimuli but with varying group labels in the IAT. Validating the correspondence between conscious and unconscious attitudes, the implicit attitude revealed stronger associations between Black + good and White + bad when athlete and politician were the labels. But when the classification dimension was changed to race (e.g., Jordan classified as Black, Helms as White), the positive implicit attitude toward the same Black athletes was eliminated. This dissociation is remarkable, given that the two attitude tasks were conducted in sequence and the explicit attitudes were strongly in favor of Black athletes and not White politicians. The cultural imposition of negativity toward the group Black interfered with the positive associations to particular favored members of the group. Mitchell et al. took these and other similar data to be evidence of the impact of the group frame in which individuals are viewed. Strong liking can turn to no liking when a shift in the construal of the attitude object is even slightly engineered.

Mitchell et al. (1998) did not take these findings to reflect opposing attitudes toward the same target; instead, they took the disjunction between the two implicit attitudes to reveal what Solomon Asch had pointed out many years ago—that the difference in attitude when performing the athlete–politician task renders a different construal of the attitude object than when performing the Black–White task. Such data lead them to view attitudes as malleable and sensitive to context and framing and suggest a departure from the assumption that attitudes come in stable, monolithic form (see Wilson, Lindsey, & Schooler, 2000, for a related but distinct view).

Malleability of Implicit Attitudes

Assumptions about the nature of implicit attitudes had led me to the conclusion that such attitudes would be hard to change, in part because the underlying associations are ones learned over a long period of time. To my great surprise, data from other labs (and now mine as well) are pointing to a contradictory assessment about

the nature of implicit attitudes. Dasgupta and Greenwald (in press) showed that when respondents are primed with photos and descriptions of admired Black individuals (e.g., Martin Luther King, Jr.) and notorious White individuals (e.g., Timothy McVeigh) they show a reduction in implicit pro-White attitude (see also Rudman, 1999, for a classroom intervention that provides a similar reduction). Blair and Ma (2000) showed that the imagining strong women leads to a reduction in the association of women as weak (relative to male as strong). Carpenter and Banaji (2000) showed the same and, in addition, that such priming effects are specific (i.e., they are restricted to change on strong–weak beliefs but do not influence good–bad attitude judgments). This line of research shows that associations between attitude object and evaluation or between belief object and attribute are more transient than previously assumed. That they can be shifted with minimal intervention is exciting both in what it reveals about the structure of attitudes and about attitude change (as perhaps not requiring deep interventions with long-lasting effects but rather as shallow interventions with shorter lasting but no less real change).

Conclusion

I have addressed, in idiosyncratic fashion, two related questions that most commonly emerge in response to research on implicit attitudes using the IAT: Are they really attitudes, and what do they predict? I suggested, through parallels with research on memory, that these questions are legitimate in some respects and inappropriate in other respects. An idea that I did not tackle is the question of implicit attitude measurement for its own sake, simply as a measure of attitude (independent of whether it relates to conscious attitudes, etc., or not, whether it is "better" or "worse" than conscious measures or not). This idea was brought home to me forcefully in an observation offered by Roddy Roediger (personal communication, April 5, 2000) that I expand on here to make the point. There is certainly some interest in knowing whether people believe they have a good memory or not, and if that question is the one of interest, one would merely ask for estimates of memory ability. But would one use a person's estimate of their memory ability when it comes to learning about memory? Or the accuracy of a person's eyesight when it comes to a decision about whether they can fly an airplane or not? Just as personal estimates would be laughable in such contexts, so too it should be with attitudes. The point is that some of these new measures are of interest simply because they allow one to firmly get away from verbal self-report measures and as such they expand the horizon of what can be learned about attitudes. In so doing, they offer a window into a mental world to which the conscious mind is not privy. Explicit attitude measures may always remain of interest given the theoretical question at hand. But implicit measures may be of interest in their own right if a measure independent of conscious feeling is sought. It may be as quaint to rely on self-reported attitude for some purposes, as it currently

would seem if one relied on people's estimates of their own intelligence as the basis of providing them admission to college.

Looking at these two areas of research on attitudes and memory has been beneficial to me, if for no reason other than to bring to the fore that differing assessments can emerge of the same underlying processes. Almost 30 years ago, researchers of implicit memory gained momentum in part through observations that amnesia patients showed intact priming but poor conscious recollection (Warrington & Weiskrantz, 1968). A decade later, Nisbett and Wilson (1977) wrote an article I regard to be among the more important contributions to psychology over the past 100 years, in which they presciently described how we, as a field, may have overestimated people's ability to introspect on the contents of their mind and the rationale for their behavior. They warned that social psychologists ought to bear in mind the dissociations between what people say and what they do, between the actual and presumed reasons for action. Although that article is well known, I do not believe its advice was taken seriously until recent years, when a focus on unconscious processes came to be of interest in its own right. Both fields of research on memory and attitude have gained enormously from breaking away from the fold of traditional measurement. The goal of this chapter was to highlight two main issues concerning what implicit attitudes are and their validity. As well, it provided an opportunity to stand on a soapbox to warn about a few matters: the strangely different standards for assessing the value of research on implicit memory versus implicit attitudes, a confusion regarding whether attitudes reveal about the individual or the culture, and the potential resistance to the idea of a lack of control over the attitudes and beliefs that are dear to the hearts of most Americans (viz., a strong sense of racial equality and egalitarianism in general).

I recall that when Bob Crowder was vastly impressed with an idea, the highest praise that he would offer in his understated way was the following: "You know, it changed my mind." So to my kindred spirit, I can simply say, "You changed my mind." But far more important than the changing of single minds lies the task, even the hope, that the work of a new generation vigorously understanding how attitudes and beliefs operate in unconscious and conscious modes will change so many minds that the arguments posed here will seem anachronistic, even defunct.

References

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Banaji, M. R., & Bhaskar, R. (1999). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, belief and brain* (pp. 139–175). Cambridge, MA: Harvard University Press.
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, 44, 1185–1193.
- Banaji, M. R., & Crowder, R. G. (1991). Some everyday thoughts on ecologically methods. *American Psychologist*, 46, 78–79.
- Banaji, M. R., & Crowder, R. G. (1994). Experimentation and its discontents. In P. E. M. & M. Gruneberg (Eds.), *Aspects of memory* (2nd ed., pp. 296–308). New York: Routledge.
- Banaji, M. R., Greenwald, A. G., & Rosier, M. (1997, October). *Implicit esteem: When collectives shape individuals*. Paper presented at the Preconference on Self, Toronto, Ontario, Canada.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 136–141.
- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgments. *Journal of Personality and Social Psychology*, 65, 272–281.
- Banaji, M. R., Lemm, K., & Carpenter, S. J. (in press). Automatic and implicit processes in social cognition. In A. Tesser & N. Schwartz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes*. London, England: Blackwell.
- Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford Press.
- Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, 62, 893–912.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, 70, 1142–1163.
- Blair, I., & Ma, J. (2000). *Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery*. Manuscript submitted for publication, University of Colorado, Boulder.
- Cameron, J. A., Alvarez, J. M., & Bargh, J. A. (2000, February). Examining the validity of implicit and explicit measures of prejudice: Is there really a bona fide pipeline? Paper presented at the first annual meeting of the Society for Personality and Social Psychology, Nashville, TN.
- Carpenter, S., & Banaji, M. R. (2000). *Implicit gender attitudes: Group membership, cultural construal, and malleability*. Unpublished manuscript, Yale University, New Haven, CT.
- Citrin, L. B., & Greenwald, A. G. (1998, April). *Measuring implicit cognition: Psychologists' and entomologists' attitudes toward insects*. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Crowder, R. G. (1989). Modularity and dissociation in memory systems. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 271–294). Hillsdale, NJ: Erlbaum.

- Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2000). *Conscious and unconscious ethnocentrism: Revisiting the ideologies of prejudice*. Manuscript in preparation, Yale University, New Haven, CT.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (in press). Psychometric properties of implicit attitude measures: Internal consistency, stability, and convergent. *Psychological Science*.
- Dasgupta, N., & Greenwald, A. G. (in press). Exposure to admired group members reduces automatic intergroup bias. *Journal of Personality and Social Psychology*.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, 36, 316-328.
- De Houwer, J. (1999). *Understanding irrelevant feature paradigms: Why is Stroop Stroop and Simon Simon?* Unpublished manuscript, University of Southampton, Southampton, England.
- Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindsay (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 269-322). New York: McGraw-Hill.
- Fazio, R. H. (in press). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*.
- Fazio, R. H., Chen, J. M., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and strength of the object-evaluation association. *Journal of Experimental Social Psychology*, 18, 339-357.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013-1027.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.
- Feynman, R. P. (1963). *Lectures in physics* (Vol. 1). Reading, MA: Addison-Wesley.
- Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 915-981). Boston: McGraw-Hill.
- Glaser, J., & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology*, 77, 669-687.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Rosier, M. (2000). Prologue to a unified theory of attitudes, stereotypes, and self-concept. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition and behavior* (pp. 308-330). New York: Cambridge University Press.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13, 141-154.
- Jacoby, L. L. (1994). Measuring recollection: Strategic vs automatic influences of association context. In C. Umiltà (Ed.), *Attention and performance*, XV (pp. 661-679). Cambridge, MA: Bradford.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 3, 306-340.
- Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology*, 10, 93-98.
- Jost, J. T., Banaji, M. R., & Greenwald, A. G. (1994, June). Experiments on (un)consciousness raising: Exploring the false fame bias in feminist samples. Paper presented at the annual meeting of the American Psychological Society, Washington, DC.
- Karpinski, A., & Hilton, J. L. (1999). *Validating attitude measures: The implicit association test as a cautionary tale*. Manuscript submitted for publication, University of Michigan, Ann Arbor.
- McGuire, W. J. (1989). The structure of individual attitudes and attitude systems. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 37-69). Hillsdale, NJ: Erlbaum.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Mitchell, J., Nosek, B., & Banaji, M. R. (1998, June). *A rose by any other name? Dissociated attitudes toward social group members*. Paper presented at the annual meeting of the American Psychological Society, Washington, DC.
- Mitchell, J., Nosek, B., & Banaji, M. R. (2000, February). *Category salience determines implicit attitudes toward Black female and White male targets*. Paper presented at the first annual meeting of the Society for Personality and Social Psychology, Nashville, TN.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226-254.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nosek, B. A. (1999, August). *Reaction time research on the Internet*. Paper presented at the 107th Annual Convention of the American Psychological Association, Boston.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2000). *Harvesting implicit group attitudes and beliefs from a demonstration website*. Manuscript under review, Yale University, New Haven, CT.
- Ochsner, K. N., & Lieberman, M. D. (2000). *The social cognitive neuroscience approach*. Unpublished manuscript, Harvard University, Cambridge, MA.
- Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (in press). Implicit attitudes and racism: The effect of work familiarity and frequency in the implicit association test. *Social Cognition*.

- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Gatenby, J. C., Funayama, E. S., Gore, J. C., & Banaji, M. R. (2000). Amygdala activation predicts performance on indirect measures of racial bias. *Journal of Cognitive Neuroscience*, 12, 729-738.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, 39, 475-543.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45, 1043-1056.
- Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-181). Amsterdam, The Netherlands: Elsevier.
- Rudman, L. A. (1999, May). Predictive utility for the implicit association test. Paper presented at the Conference on Construct Validity in Implicit Social Cognition, National Science Foundation, Zion, IL.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & McGhee, D. E. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the implicit association test. *Social Cognition*, 17, 1-29.
- Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, 26, 1315-1328.
- Salovey, P., Mayer, J. D., & Caruso, D. (in press). The positive psychology of emotional intelligence. In C. R. Snyder & S. J. Lopez (Eds.), *The handbook of positive psychology*. New York: Oxford University Press.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 501-518.
- Sears, F. (1980). *University physics* (5th ed.). Reading, MA: Addison-Wesley. (Original work published 1902)
- Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (in press). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion*.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Wallace, S. R. (1965). Criteria for what? *American Psychologist*, 20, 411-417.
- Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972-974.
- Weitz, J. (1961). Criteria for criteria. *American Psychologist*, 16, 228-231.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.

Analys

Bennet Mur

As a Yale give a few in 1947, and the 12 c year before) ε like the Holy proceedings l around the ta the very last was very diff almost got up

At the tin experimental l rats 20 hours Jaynes was sle all of us were what I had c approach-av proach-avoida have survived

Not only w Could anyone for attention? I informal course verbal learners he knew nothi

This work was : Research Counc helpful commen program for Fig