Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT

Michael A. Olson and Russell H. Fazio
Ohio State University

The authors argue that the Implicit Association Test (IAT; A. G. Greenwald, D. E. McGhee, & J. L. K. Schwartz, 1998) can be contaminated by associations that do not contribute to one's evaluation of an attitude object and thus do not become activated when one encounters the object but that are nevertheless available in memory. The authors propose a variant of the IAT that reduces the contamination of these "extrapersonal associations." Consistent with the notion that the traditional version of the IAT is affected by society's negative portrayal of minority groups, the "personalized" IAT revealed relatively less racial prejudice among Whites in Experiments 1 and 2. In Experiments 3 and 4, the personalized IAT correlated more strongly with explicit measures of attitudes and behavioral intentions than did the traditional IAT. The feasibility of disentangling personal and extrapersonal associations is discussed.

Implicit measures have enjoyed widespread use in social psychology in recent years. The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) has become a particularly popular implicit lens for viewing such social phenomena as prejudice (e.g., McConnell & Liebold, 2001), self-esteem (e.g., Greenwald & Farnham, 2000), and social identity (e.g., Greenwald, Banaji, Rudman, Farnham, & Nosek, 2002; for reviews, see Fazio & Olson, 2003b; Greenwald & Nosek, 2001). Research explosions concerning a given topic occur periodically, but such a surge of research based on a particular measurement *tool* rarely has been seen in the field. Because research that uses the IAT as a lens to view the human mind is poised to have a tremendous impact on the shape of social psychology, it is important to understand the mechanism underlying its function. The research presented here addresses this issue.

By their definition, lenses are prone to alter the image they present, and the IAT may be no exception. Accordingly, the research presented here also addresses the question of whether the IAT tends to distort what it reveals about the workings of the mind. Specifically, we propose that the IAT is contaminated by *extrapersonal associations*—associations that are available in memory but are irrelevant to the perceived likelihood of personally experiencing a positive or negative outcome on interaction with the attitude object (for a related view, see Karpinski & Hilton, 2001). Moreover, we present a method of solving this contamination problem.

A Brief History of the IAT

The IAT is said to measure associations between constructs by forcing participants to associate them with the same response keys (Greenwald & Nosek, 2001). In an IAT measuring attitudes toward Blacks, for example, participants are presented with pleasant and unpleasant words (e.g., *love* and *bombs*) and typically Black and White names (e.g., *Tyrone* and *Hank*) or faces. They categorize these four classes of items by pressing one of two response keys, forcing them sometimes to associate Black and pleasant (and, hence, White and unpleasant) and sometimes to associate Black and unpleasant (and, hence, White and pleasant). It is argued that this test reveals racial prejudice to the extent that participants find it more difficult (as assessed by either response latencies or categorization errors) to make the former relative to the latter categorization.

The IAT has been adapted to assess the strength of the association between a host of important psychological variables, such as the self with valence (i.e., self-esteem), gender, health beliefs, and various personality traits (e.g., Greenwald, Banaji, et al., 2002; Jordan, Spencer, & Zanna, 2002; Marsh, Johnson, & Scott-Sheldon, 2001; Nosek, Banaji, & Greenwald, 2002b; Rudman, Greenwald, & McGhee, 2001); in-groups and out-groups with various traits and evaluations (e.g., Ashburn-Nardo, Voils, & Monteith, 2001; Blair, Ma, & Lenton; 2001; Florack, Scarabis, & Bless, 2001; Lowery, Hardin, & Sinclair, 2001); and consumer products with various attributes (e.g., Maison, Greenwald, & Bruin, 2001). In short, the IAT has quickly become a

Portions of this research were presented at the Annual Meeting of the Midwestern Psychological Association, Chicago, May 2003. This research was supported by Senior Scientist Award MH01646 and Grant MH38832 from the National Institute of Mental Health. We thank Barbara Fazio and Suzanne Miller for their thoughtful contributions to some of the ideas presented in this paper and Julie Brown, Lindsey Cudlike, Roxanne Flowers, Alexis Greenhut, Kevin Griffin, Katie Hershey, Don Ladwig, Nicole DeLand, Sarah Lifland, and Dorey Meslemani for help in data collection; Brian Nosek for feedback on an earlier version of this article; and Tony Greenwald for providing both feedback and stimulus materials.

Correspondence concerning this article should be addressed to Michael A. Olson or Russell H. Fazio, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210-1222. E-mail: olson.127@osu.edu or fazio@psy.ohio-state.edu

Journal of Personality and Social Psychology, 2004, Vol. 86, No. 5, 653–667
 0022-3514/04/$12.00 DOI: 10.1037/0022-3514.86.5.653

preferred implicit measurement tool for many important social–cognitive and personality variables.

Evidence demonstrating the value of the IAT as a measurement technique has accumulated quickly in its brief history. Research documenting its discriminant validity often has used a known-groups approach. For example, an in-group preference has been found for several different pairs of naturally occurring groups (e.g., Japanese and Korean Americans, Jews and Christians, East and West Germans, and smokers and nonsmokers; Greenwald et al., 1998; Kühnen et al., 2001; Swanson, Rudman, & Greenwald, 2001). IAT effects also have been found to relate to particular states of amygdala activation (Phelps et al., 2000), and the measure has proven sensitive to detecting experimentally created preferences for in-groups (Ashburn-Nardo et al., 2001; see also Greenwald, Pickrell, & Farnham, 2002) and newly conditioned attitudes toward novel objects (Olson & Fazio, 2001). Regarding predictive validity, McConnell and Liebold (2001) demonstrated correspondence between a racial attitude IAT and certain nonverbal behaviors like speaking time, smiling, and speech errors while interacting with a Black relative to a White experimenter. Self-esteem IAT scores have been found to relate to reactions to success-versus-failure feedback (Greenwald & Farnham, 2000) and persistence in the face of failure (Jordan et al., 2002). Very recent work has begun using the IAT to assess the interrelationships between the self, attitudes, and group memberships (Greenwald, Banaji, et al., 2002; for reviews of IAT research, see Fazio & Olson, 2003b; Greenwald & Nosek, 2001).

Possible Mechanisms Underlying the IAT

The research cited above provides merely a sample of the prolific use that the IAT has seen. In this work, it has been assumed that the IAT measures what it purports to measure—associations between attributes (e.g., "pleasant") and concept categories (e.g., "Blacks"). According to its developers, the IAT works because "if two concepts are highly associated, the IAT's sorting tasks will be easier when the associated concepts share the same response than when they require different responses" (Greenwald & Nosek, 2001, p. 85). Recently, increasing attention has been paid to more fundamental issues: the nature of the associations revealed by the IAT and the mechanism underlying the IAT effect. Several models designed to describe the workings of the IAT have been proposed (e.g., Brendl, Markman, & Messner, 2001; Mierke & Klauer, 2001; Rothermund & Wentura, 2001), but De Houwer's (2001) model is most relevant for the present purposes, and we briefly review it next.

According to De Houwer (2001), IAT respondents attend only to the features of items relevant to making the required discrimination, that is, the items' category membership. For example, in an IAT involving flowers and insects, respondents might be presented with *rose*, which they would identify as a flower by pressing a corresponding button. What matters most to respondents, according to De Houwer, is the membership of *rose* in the category "flower," for their task involves only discriminating between flowers and insects. Thus, exemplars will be attended to only to the point where their category membership is derived, and the specific exemplars used to represent the category should make little difference (as long as they are clearly members of the category and do not themselves affect the meaning of the category; see Govan & Williams, in press). In short, evaluative associations involving individual exemplars of the categories should have little impact on the IAT compared with associations involving the category labels.

To test this hypothesis, De Houwer (2001) conducted a British–foreigner IAT (with British participants), using both positive and negative Brits (e.g., the Queen Mother, a mass murderer) and foreigners (Einstein, Hitler), and found that Brits were more easily associated with "pleasant" than were foreigners, regardless of the individual British and foreign items. In other words, Einstein and Hitler were both more easily associated with unpleasant items because both were "foreigners." Thus, the instructions to categorize the items force them to be construed only as representatives of their respective categories, resulting in IAT scores based predominantly on associations to category labels (see also Govan & Williams, in press; Mitchell, Nosek, & Banaji, 2003; Olson & Fazio, 2003). So it seems that the IAT has relatively little to do with the evaluation associated with the individual exemplars. It is doubtful, for example, that respondents experience the automatic activation of positivity in response to the name of a well-known mass murderer or negativity in response to Einstein except when encouraged to categorize such exemplars as fellow Brits versus foreigners.

Contamination of the IAT

The IAT's operation at the level of the category label, instead of the individual exemplar, suggests that researchers should consider the category labels and not the individual exemplars to be the objects most directly relevant to the IAT. Moreover, if the IAT has little to do with the automatic activation of evaluations in response to the exemplars, then it is assessing only the ease with which a respondent can associate a given category label with a given valence. However, any such ease (or difficulty) need not reflect solely the influence of the individual's attitude. Instead, it, and thus performance on the IAT, may be influenced by associations to the category that are unrelated to the individual's evaluation of the category. However, before presenting our own position, we review some related work that to varying degrees suggests that the IAT may be contaminated by attitude-irrelevant associations.

Perhaps most directly pertinent is work by Karpinski and Hilton (2001), who demonstrated a dissociation between explicit measures of attitudes and the IAT. Specifically, an IAT designed to assess preferences for apples versus candy bars showed no correlation with explicit measures of the same construct. Karpinski and Hilton concluded that the IAT was contaminated by what they called *environmental associations*—culturally shared but not necessarily individually accepted positive information about apples (and negative information about candy bars).

It is important to acknowledge that dissociations between explicit and implicit measures may be caused by many variables. Certainly, individuals' concerns with social desirability when responding to explicit measures can produce such dissociations (for reviews, see Blair, 2001; Dovidio, Kawakami, & Beach, 2001; Fazio & Olson, 2003b). However, in the case highlighted by Karpinski and Hilton (2001), there appears little reason to suspect that people are unwilling to honestly report their attitudes toward apples and candy bars. Another possible explanation, that people are unaware of their attitudes toward candy bars and apples and that the IAT is able to tap into these unconscious attitudes in a way

explicit measures cannot, lacks both intuitive appeal and parsimony. Peoples' vast experience with these attitude objects is reason alone to doubt that awareness of their attitudes is the problem. An even more remote possibility is that the explicit measures were the problem—that they were somehow not tapping into participants' evaluations of apples and candy bars. However, one would be hard-pressed to dispute the face validity of a question like "how much do you like apples?" to tap into one's evaluation of apples.[1] Finally, the IAT and explicit measures may be tapping two different attitudes (or two different components of the same attitude) toward the same object (Wilson, Lindsey, & Schooler, 2000). For example, one might argue that the IAT is more attuned to affective associations and that explicit measures tap semantic aspects of evaluations (or vice versa). However, this approach risks rendering the problem a semantic one, where each time a dissociation is observed, a different component of the attitude must be posited. According to this view, the attitude becomes a product of the method of measurement—and the perceiver is forced to find a place for as many attitude components as dissociated measures of them. Clearly, some theoretical reason for supposing that different measures will show different patterns of results is needed before all of the measures are thrown into the matrix.

In providing evidence for their environmental associations view, Karpinski and Hilton (2001) highlighted not only that the IAT and explicit measures of attitudes did not correlate but also that participants displayed a more marked preference for apples over candy bars on the IAT as compared with the explicit measures. They pointed out that society portrays apples quite positively but is more ambivalent about candy bars and that this pattern was reflected in their participants' IAT scores. Karpinski and Hilton also found that the apple–candy bar IAT was unable to predict behavior. That is, participants' IAT-derived preferences did not relate to their behavior when they were offered a choice between an apple and a candy bar at the end of the experiment, although explicit measures did.

Additional findings in the literature on implicit measures suggest a similar point. First, a larger proportion of Whites appears negative toward Blacks on the IAT compared with other implicit measures of racial prejudice. For example, about 80% of White participants have shown some degree of negativity toward Blacks on the IAT (e.g., Nosek, Banaji, & Greenwald, 2002a) compared with about 50% on priming measures (e.g., Fazio, Jackson, Dunton, & Williams, 1995). Given Blacks' negative portrayal by much of the media, even people for whom positivity is automatically activated in response to Blacks ought to have readily available in memory a host of negative associations with Blacks. This would inflate estimates of prejudice on the IAT if the IAT is contaminated by this general knowledge. Relatedly, and surprisingly, Blacks do not show an in-group preference on the IAT (Nosek et al., 2002a), a pattern that stands in stark contrast to what has been repeatedly observed on both implicit and explicit measures (Blair, 2001; Dovidio et al., 2001; Fazio et al., 1995). Yet, Black respondents to the IAT clearly are knowledgeable regarding the generally negative portrayal of Blacks, and this knowledge may facilitate their associating Blacks and negativity on the IAT.

In the self domain, Greenwald and Farnham (2000) observed that women more easily associated themselves with traditional female stereotypes on the IAT than on explicit measures. This may be due to either a reluctance to admit the extent to which traditional gender roles make up one's identity or a lack of awareness of the extent to which society's stereotypes about the sexes influence one's identity implicitly. However, the pattern also is consistent with the possibility that women can easily recall instances of their own gender's association with certain traditional female stereotypic roles despite their personal beliefs and that this information can facilitate associating the self with traditional female roles on the incompatible trials.

Personal Versus Extrapersonal Associations

In sum, several pieces of evidence, some more direct than others, converge to suggest that the IAT may be contaminated by what we refer to as *extrapersonal associations.* Because the personal–extrapersonal distinction is critical to our argument and because our argument differs somewhat from what has been advanced in previous work, it is important that we elaborate on the distinction. Our reasoning regarding personal associations rests squarely on the view of attitudes as associations in memory between the attitude object and one's summary evaluation of the object (Fazio, 1995; Fazio, Chen, McDonel, & Sherman, 1982; Fazio, Powell, & Herr, 1983). Thus, we begin with the premise that an attitude is by definition an inherently personal association. In addition, and consistent with a long-standing perspective among attitude theorists (e.g., Fazio, 1995; Katz, 1960; Smith, Bruner, & White, 1956), we maintain that the functional utility of attitudes lies in their directing attention, categorization, and ultimately behavior in a manner that maximizes the likelihood of the individual's experiencing positive outcomes and avoiding negative ones (for a review of relevant evidence regarding the object-appraisal function of attitudes, see Fazio, 2000). In gauging the appropriateness of approach or avoidance behavior, then, what proves functional for individuals is to have their own attitudes automatically activated on their encountering the attitude object, that is, their personal evaluative associations regarding the object.

Just as one would expect from this theoretical perspective, considerable priming research aimed at assessing automatic attitude activation has found that participants' idiosyncratic attitudes are activated in response to attitude objects when they are presented as primes. It appears to be neither the valence that a majority of people associate with the primed object nor the view implied by some culturally shared perspective that is automatically activated, but rather the participants' own personal attitudes. For example, individuals whose evaluations of an object differ from popular opinion show activation of attitudes consistent with their own previously reported attitudes and behavior instead of the general consensus (e.g., Fazio, 1993; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Sherman, Presson, Chassin, Rose, & Koch, 2002). Moreover, estimates of these idiosyncratic attitudes based on responses to the attitude objects when presented as primes have been shown to predict behavior toward the object (e.g., Bessenoff & Sherman, 2000; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995).

[1] Although it is also likely that simple measurement error weakens correlations between two measures (Cunningham, Preacher, & Banaji, 2001), it probably cannot account (at least fully) for correlations of zero.

This is not to say that personal associations may not have been influenced by information transmitted by specific others or by the culture in general (for further discussion of this issue, see Banaji, 2001; Lowery et al., 2001; Olson & Fazio, 2003). One of the primary lessons of social psychology is that the individual is greatly affected by socialization experiences. Cross-cultural research has amassed a host of examples of how individuals' attitudes, beliefs, and norms are a product of the culture in which the individual is socialized (e.g., Fiske, Kitayama, Markus, & Nisbett, 1998). In fact, some researchers have argued that because people are all, at least in part, products of the culture in which they live, it makes little sense to attempt a separation between what is "personal" and what is "cultural" (Banaji, 2001).

Although it is indeed difficult to conceive of an individual without social or cultural influence, the separation we propose is far less profound. Our argument is merely that individuals' attitudes can vary from the cultural norm or from what they know to be the evaluations of specific others, as is probably the case for people with allergies to peanuts or a penchant for pickled herring. For these individuals, it is likely that information that is opposite in valence from their attitudes is available in memory. Information of this sort, that is, information that does not contribute to an individual's personal evaluation, is what we refer to as extrapersonal in nature. Such extrapersonal associations are irrelevant to the anticipated likelihood of one's experiencing a positive or negative outcome on interacting with the attitude object. Yet, this information is available in memory and, as we argue below, can influence participants' IAT performance.

It is important to note, however, that we do not make the additional assumption that the personal–extrapersonal distinction necessarily corresponds with individuals' acceptance or endorsement of the association, as Karpinski and Hilton (2001) argued.[2] Our position is that attitudes may form on the basis of both personally accepted information and information to which one never deliberately acquiesced. We do not question that individuals may possess attitudes that, when met with their attention, they prefer to correct for or disavow in some way (Dunton & Fazio, 1997; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Plant & Devine, 2001; Wegener & Petty, 1995; Wilson & Brekke, 1994).

Let us now turn to the issue of how extrapersonal associations might affect IAT performance. Our reasoning is premised on the notion that participants are motivated to follow the IAT task instructions to respond as quickly as possible in categorizing the items presented. Participants are likely to be influenced by whatever information is available to them in memory when mapping two concepts onto the same response key. Our claim is that some of this information may affect response latencies to the mapping task even though it does not contribute to one's attitude. Consider, for example, an IAT designed to measure one's attitude toward peanuts. Most people feel at least some degree of positivity toward peanuts, and therefore would probably be relatively quick to map pleasant items onto the same response key as "peanut," because their attitude serves as a basis for easily associating "peanut" with "pleasant." However, participants may also recall other information about peanuts that has positive implications, even though it does not provide any basis for their evaluations. They may, for example, remember that George Washington Carver was able to invent hundreds of uses for the peanut, or that the Planters brand peanut mascot wears a rather classy top hat. This information may not have been activated had it not been for the specific nature of the IAT instructions and the task-induced goal to associate peanuts with positivity. The consequence of recalling this information, however, would be the facilitation of associating peanuts with "pleasant" and a more positive attitude estimate from the IAT.

It also may be possible for the IAT to reflect the influence of an extrapersonal association that is opposite to one's actual attitude. Counterattitudinal extrapersonal information would be available to nearly anyone who disliked peanuts but was socialized in a "pro-peanut" culture, including someone with a severe allergy to (and, hence, a strongly negative attitude toward) peanuts. When attempting to map "peanuts" and "pleasant" onto the same response key, positive extrapersonal information about peanuts of the sort noted above should prove useful. Activation of such information, whether it be strategically or unconsciously, would facilitate associating "peanuts" with "pleasant." It appears at least theoretically possible, then, that the wealth of positive extrapersonal information available in memory might make this individual appear relatively more positive in his or her evaluation of peanuts than he or she otherwise would, even though the very sight of peanuts may promote a gag reflex. In sum, because of the explicit demand on participants to associate pleasant and unpleasant items and attitude objects, the IAT task may prompt the recollection of attitude-irrelevant information that is available in memory, resulting in an attitude estimate that is contaminated by these extrapersonal associations.

Decontamination of the IAT

Three features of the IAT as it is typically administered may encourage the use of extrapersonal associations. First, consider the category labels themselves. The labels "Pleasant" and "Unpleasant" carry a specifically normative implication. That is, there is something about the item being presented that makes it a member of the category "pleasant" or "unpleasant," not something about the participant's attitude toward it. Second, the pleasant and unpleasant items typically presented as members of the attribute categories "pleasant" and "unpleasant" are universally pleasant (e.g., *love*) and unpleasant (e.g., *bombs*). The items are typically portrayed as either normatively positive or negative—virtually everyone either likes them or dislikes them. Third, the typical IAT includes the provision of feedback when the participant has made an error. This feedback certainly suggests that there is a normatively correct response. These factors might increase the accessibility of normative information relevant to solving the mapping problem posed by the IAT, leaving attitudes as only one of several potential types of associations that influence performance on the mapping tasks.

The present experiments tested various instantiations of this reasoning. Changing the labels of the pleasant and unpleasant categories to something less likely to be construed normatively might decrease the influence of normative information when solv-

[2] Indeed, we have repeatedly demonstrated in other research that some individuals do accept their attitudes as legitimate and a proper basis for behavior, whereas others do not endorse them as such. For example, we have demonstrated that some White individuals with negative attitudes toward Blacks attempt to correct for the influence of their attitudes (e.g., Olson & Fazio, in press; Towles-Schwen & Fazio, 2003).

ing the IAT's mapping task (Experiments 1–4). We chose the labels "I like" and "I don't like" for the experiments reported here. Moreover, we identified items for the pleasant–unpleasant categorization task that, although attitude evoking, are not normatively associated with a given valence (Experiments 1 and 3 only). That is, we chose evaluation-laden items for which there is little social consensus among college students (e.g., *coffee, football*). Finally, we did not include error feedback in our modified IAT (Experiments 1–4). In other words, we took various steps to personalize the IAT. All the experiments reported here involved a comparison of a personalized IAT with a traditional IAT.

In Experiment 1, participants completed either a traditional racial attitude IAT or our personalized version. Given the assumption that Blacks are portrayed relatively negatively, we predicted that the traditional IAT would reveal more prejudice than the personalized IAT.

Experiment 1

Method

Participants. Seventy-eight non-Black undergraduates at a Midwestern university participated for course credit. Five participants were excluded for committing a large number (> 20%) of errors, resulting in a final sample of 49 women and 24 men.

Materials and procedure. The IAT was introduced as a "categorization task," and the instructions informed participants that they would be categorizing a variety of items that would appear on the computer screen. Procedures were modeled closely after Greenwald et al. (1998). There were 12 total blocks, each consisting of 50 trials. On a given trial, a stimulus word was presented in the center of the screen, and the participant's task was to categorize it by pressing a corresponding button on the keyboard. Each block was preceded by a set of instructions presented on the screen that informed participants of the type of items that they would be categorizing as well as the meaning of the keys (key labels remained on the screen throughout each block). The first 2 blocks provided participants with practice categorizing Black (e.g., *Theo*) and White (e.g., *Chip*) names. The next 2 blocks consisted of trials requiring the categorization of valenced words. In the traditional version, the task was described as one involving the discrimination of "pleasant and unpleasant items," and the labels "Pleasant" and "Unpleasant" appeared on the screen to identify the meaning of the keys. The items presented (taken from Greenwald et al., 1998) were very clearly and universally pleasant (e.g., *freedom*) and unpleasant (e.g., *murder*; the Appendix lists all stimulus items). In the personalized version, the discrimination task was described as involving "things you might like or dislike," the keys were labeled "I like" and "I don't like," and the items presented were pretested as having no clear normative evaluation (i.e., a mean that did not differ from zero) but a large degree of variability in personal evaluations (i.e., a relatively large standard deviation; e.g., *coffee, football*). The final difference between the two versions of the pleasant–unpleasant practice blocks was that in the traditional version (where there was a normatively correct response), errors were followed by a red *X* presented on the screen, which disappeared after the correct response was made. No error feedback was presented in the personalized version.

Blocks 5–7 were critical combined blocks, where one race was paired with the positive category and the other was paired with the negative (depending on the counterbalancing condition to which each participant was randomly assigned). As before, the traditional version involved the presentation of normatively valenced pleasant and unpleasant items (with the labels "Pleasant" and "Unpleasant"), and the personalized version involved the presentation of items with no normative evaluation (with the labels "I like" and "I don't like"). Blocks 8 and 9 were practice blocks involving the categorization of Black and White names and were identical for the two conditions. Blocks 10–12 were also critical combined blocks and were identical to Blocks 5–7, but the race that was paired with the positive category was now paired with the negative (and vice versa). For convenience, we refer to the blocks in which Blacks shared a response key with the positive category (and, hence, Whites with the negative) as the *Black*/+ blocks. Likewise, *Black*/− refers to blocks during which Blacks and the negative category (and, hence, Whites and the positive category) shared a response key.

Participants then completed a second IAT identical to the first, only under different instructions; those who had completed the traditional IAT now completed a personalized IAT (and vice versa). All participants completed two more practice blocks involving either "pleasant" and "unpleasant" or "I like" and "I don't like." These additional practice blocks were designed to help participants redefine the meaning of the pleasant/I like and unpleasant/I don't like task. After two more blocks involving Black and White names, participants completed three critical combined blocks, two more practice blocks involving names, and three critical combined blocks with the races now switched in their affiliation with positive and negative (again, with the order of Black/+ and Black/− blocks counterbalanced). After completing the IAT, participants were debriefed, thanked, and dismissed.

Results and Discussion

Following standard practice (Greenwald et al., 1998), response latencies from the first two trials from each of the Black/− and Black/+ blocks were dropped, and the remaining trials were natural log transformed. All analyses were conducted using these transformed IAT difference scores, but means are reported in terms of raw latency differences for ease of exposition. Means for each block type were then computed and were entered into a 2 (block type: Black/− vs. Black/+) × 2 (IAT type: traditional vs. personalized) × 2 (order: traditional first vs. personalized first) analysis of variance (ANOVA), with repeated measures on the first two factors.

The typical IAT effect of prejudice against Blacks was observed in a main effect of block type, $F(1, 71) = 36.17$, $p < .01$, such that participants were slower to respond in the Black/+ ($M = 859.53$, $SD = 146.74$) relative to the Black/− ($M = 711.89$, $SD = 115.53$) blocks. There was also a large effect of IAT type, $F(1, 71) = 176.32$, $p < .01$, such that regardless of block type, participants responded more slowly while performing the personalized IAT ($M = 819.21$, $SD = 143.41$) relative to the traditional IAT ($M = 751.27$, $SD = 127.76$). However, these effects were qualified by the critical two-way Block Type × IAT Type interaction, indicating that the two versions of the IAT revealed different levels of prejudice, $F(1, 71) = 3.98$, $p = .05$.[3] On the traditional IAT, the mean of response latencies on the Black/+ task was 832.85 ($SD = 157.11$) compared with 669.68 ($SD = 123.88$) on the Black/− task. On the personalized IAT, the Black/+ task mean was 885.25

[3] The three-way interaction was significant, $F(1, 71) = 7.93$, $p < .01$. However, interaction effects involving order merely reflect a well-documented effect of practice on the IAT in that participants show less prejudice on the second IAT that they perform (e.g., Wittenbrink, Judd, & Park, 2001). Less prejudiced IAT scores were produced on the second compared with the first IAT whether participants completed the traditional ($M = 181.10$, $SD = 114.76$) and then the modified ($M = 111.72$, $SD = 101.30$) IAT or the modified ($M = 154.77$, $SD = 106.09$) and then the traditional ($M = 142.60$, $SD = 127.39$) IAT. It is important to note that less prejudiced IAT scores were found on the modified version regardless of task order.

(SD = 167.58) compared with 753.17 (SD = 136.29) on the Black/− task. Stated another way, in terms of typical IAT scores (computed by subtracting response latencies on the compatible blocks from those on the incompatible blocks), the personalized version produced less racial bias on average (M = 131.6, SD = 105.8) than did the traditional version (M = 163.17, SD = 121.5), $t(72) = 2.10$, $p < .05$. Attitude estimates from the two versions of the IAT were moderately correlated, $r(73) = .37$, $p < .01$.

Well after Experiment 1 was conducted, Greenwald, Nosek, and Banaji (2003) suggested a new scoring algorithm for the IAT based on analyses of large samples of data collected via the Internet. They recommended using untransformed latencies, omitting trials with responses greater than 10,000 ms, omitting participants whose response latencies were shorter than 300 ms on over 10% of the trials, replacing the latencies of trials on which a categorization error was committed with the mean of the block during which the error was committed plus a 600-ms "penalty," and computing a pooled standard deviation from the first pair of blocks involving the combined categorization task and another for each subsequent block pair. For each block pair, the difference between the mean latencies from the block involving a given response mapping and the block involving the reverse mapping is computed. This difference is then divided by the pooled standard deviation of those two blocks, resulting in a D ratio. The average of these D scores from each of the incompatible–compatible block pairs serves as the attitude index in this algorithm.

For exploratory purposes, we computed IAT scores based on the Greenwald et al. (2003) recommendations. We implemented all aspects of the scoring algorithm with two exceptions. First, because errors are impossible to detect on attribute trials of the personalized IAT, we did not implement the prescribed error penalty for either version of the IAT. This was necessary to maintain consistent treatment of data from the two experimental conditions. Second, we felt uncomfortable following the prescription to compute IAT scores on the basis of raw latencies; the distributions of the IAT latencies were highly skewed, and averaging raw latencies effectively assigns disproportionate weight to trials characterized by slow responses. Thus, in keeping with long-standing convention in response-time research, transformed latencies were used instead. Using this algorithm, the personalized IAT reflected less negativity toward Blacks (M = .36, SD = .28) than did the traditional version (M = .48, SD = .29), $t(71) = 3.43$, $p < .01$, mirroring the results reported above.[4]

In order to assess the possibility that a change in participants' accuracy affected attitude indices, accuracy rates from critical trials involving a Black or White categorization were examined from each IAT. On average, participants responded correctly on 93% of these trials in the traditional IAT and on 92% of these trials in the personalized IAT ($t < 1$). Thus, participants' accuracy motivation was equivalent for both versions of the task. Accuracy for the attribute items cannot be compared similarly, because these items differed for the two versions of the IAT, and the items used in the personalized version were not consensually positive or negative. However, considering pleasant–unpleasant trials on critical blocks in the traditional IAT on average, the expected 50% of items were categorized as "pleasant." In the personalized IAT, 56% of these items were categorized as "liked." Thus, participants did not appear to be using a "like everything" or "dislike everything" heuristic on the personalized IAT.

In sum, our prediction that the personalized version would decrease the amount of racial prejudice revealed by the IAT was confirmed.

Experiment 2

Our attempt to personalize the IAT in Experiment 1 was successful in the sense that there was a significant reduction in prejudice exhibited on the personalized version of the IAT relative to the traditional version. Experiment 2 was an attempt to replicate the effect, with two important modifications. First, to avoid any task-order complications (see Footnote 3), the experiment involved a completely between-subjects design; participants were randomly assigned to either the traditional or the personalized versions. Second, the two versions of the IAT used identical stimuli. As we argue above, De Houwer's (2001) analysis suggests that the construal of the items presented in the IAT is largely a function of the category labels. Therefore, changing the labels from "[Un]pleasant" to "I [don't] like" may be enough to direct participants to construe the items presented in terms of their own attitudes and to reduce the influence of extrapersonal associations used to solve the IAT's mapping problem. In Experiment 2, then, the only difference between the two versions involved the category labels and the presence or absence of error feedback. The traditional version's normatively pleasant and unpleasant items were used for both conditions.

Method

Participants. As in Experiment 1, participants were 114 non-Black undergraduates at a medium-sized Midwestern university who participated for course credit. Five participants were omitted because of high error rates (> 20%). The final sample consisted of 64 women and 45 men.

Materials and procedure. These were identical to Experiment 1 with the following exceptions. The normatively pleasant and unpleasant items used in the traditional condition of Experiment 1 were used for both the traditional and the personalized conditions. Participants were assigned to either the traditional or the personalized condition and performed only one version of the task. However, male and female names of each race were separated, and participants completed three blocks each of the Black/+ and Black/− tasks for both genders. Thus, the number of compatible and incompatible blocks completed overall was double that of Experiment 1 (just as in Greenwald et al., 1998). The order in which participants performed the Black/+ versus Black/− blocks was counterbalanced, as was the order in which male versus female names were used. After completing all of the IAT blocks, participants were debriefed, thanked, and dismissed.

Results and Discussion

Mean response latencies from the Black/+ and Black/− blocks were calculated as in Experiment 1 and were entered into a 2 (block type: Black/− vs. Black/+) × 2 (IAT type: traditional vs. personalized) ANOVA, with repeated measures on the first factor. A block type effect indicated a large general prejudice effect, $F(1, 108) = 319.89$, $p < .01$, such that participants were slower to

[4] The loss of one degree of freedom compared with earlier analyses was due to the omission of a participant because of excessively fast responding (response latencies on more than 10% of trials were less than 300 ms), in accordance with Greenwald et al.'s (2003) algorithm.

respond in the Black/+ ($M = 901.89$, $SD = 170.67$) relative to the Black/− ($M = 714.12$, $SD = 128.97$) blocks. However, this effect was qualified by an interaction involving block type and IAT type, $F(1, 108) = 4.07$, $p < .05$. On the traditional IAT, the mean of response latencies on the Black/+ task was 930.63 ($SD = 173.99$) compared with 721.24 ($SD = 135.69$) on the Black/− task. On the personalized IAT, the Black/+ task mean was 874.18 ($SD = 164.21$) compared with 707.25 ($SD = 122.99$) on the Black/− task. Thus, the personalized version of the IAT showed less prejudice ($M = 166.93$, $SD = 116.73$) than the traditional version ($M = 209.39$, $SD = 103.24$), $t(108) = 2.02$, $p < .05$. Using the variation of Greenwald et al.'s (2003) scoring algorithm described in Experiment 1, the personalized IAT still reflected less prejudice ($M = .46$, $SD = .29$) than the traditional version ($M = .51$, $SD = .22$), though not significantly so, $t(108) = 1.07$, $p = .28$.

It is important to note that the earlier ANOVA revealed no main effect of IAT type ($F < 1.5$, $p > .20$), indicating that response latencies were no slower or faster in general depending on which version of the IAT participants completed. Also, accuracy rates on critical trials involving the Black–White discrimination were roughly equal for both versions of the IAT (94%, $t < 1$). A similar pattern was apparent on the attribute trials—accuracy rates were 94% and 93% for the personalized and traditional IATs, respectively ($t < 1$).

In Experiment 1, three modifications were made to the traditional procedure to personalize the IAT: (a) the labels "Pleasant" and "Unpleasant" were changed to "I like" and "I don't like," (b) less normatively valenced stimulus items were used, and (c) error feedback was not provided. Attitude estimates were less indicative of racial prejudice on this personalized IAT. In Experiment 2, this reduction in prejudice was found even when the only changes were in the category labels and whether or not participants received error feedback. This pattern of results is consistent with our view that personalizing the IAT by prompting participants to focus more on their own evaluations and less on normative information results in less negative attitude estimates. Arguably, this is because there is a large amount of negative information about Blacks in the greater society—information that can be utilized more readily in the traditional version of the IAT to help solve the mapping problem the IAT presents.

The results from Experiment 2 also allow us to rule out a potential alternative explanation. Recall that in Experiment 1, response latencies were generally slower on the personalized IAT. One might argue that the reduction in prejudice observed on this version of the IAT was a result of this generalized slowing effect—that perhaps more cautious responding allowed for a reduction in the influence of prejudicial associations. However, Experiment 2 revealed no effect of IAT type on overall response latencies. Participants who completed the personalized version of the IAT were no slower than those completing the traditional version, even though they showed less prejudice. Thus, the slowing of response latencies observed on the personalized IAT in Experiment 1 apparently was due to the use of nonnormatively valenced positive and negative items. Most important, this slowing effect cannot explain the differential prejudice revealed by the two versions of the IAT in Experiment 2. Similarly, and as in Experiment 1, participants were no less accurate in the personalized IAT, which rules out the possibility that a shift in response thresholds produced the prejudice-reduction effect.

It also is worth noting the simplicity and obvious face validity of the experimental manipulation. Because it is difficult to discern whether any given piece of information has contributed to the attitude or not, the personal–extrapersonal distinction is inevitably characterized by some conceptual fuzziness. What is extrapersonal information for one individual can form the very basis for another individual's attitude. Operationally, however, the distinction is clear. The "I like" and "I don't like" labels necessarily invoke personal associations, whereas the more traditional labels of "Pleasant" and "Unpleasant" allow for the possibility of considering extrapersonal associations as well as personal ones.

In Experiment 3, we sought more direct evidence for the value of personalizing the IAT by focusing on a domain in which correspondence with explicit measures seems both likely and appropriate. Although it is certainly socially desirable to present oneself as healthy, and hence, individuals may be more likely to claim to like apples relative to candy bars, they are probably somewhat less motivated to behave in a socially desirable fashion here than when completing a direct, explicit measure of racial attitudes. Recall that Karpinski and Hilton (2001) observed null relations between an IAT assessing preferences for apples versus candy bars and explicit measures of attitudes toward the same objects. If the personalized version of the IAT removes some of the contamination of extrapersonal associations, then we should expect it to correlate better with explicit measures and behavioral intentions in this relatively less socially sensitive domain. We tested this hypothesis in Experiment 3.

Experiment 3

We have argued that Blacks are portrayed relatively negatively by society and that this information can be used in a way that makes people appear relatively prejudiced on the IAT. Similarly, Karpinski and Hilton (2001) reported an apples–candy bar IAT where participants appeared far more positive toward apples than explicit measures and actual choice behavior indicated. In Experiment 3, we again tested the extrapersonal association hypothesis, but because of apples' relatively positive portrayal, we predicted that a traditional IAT would show positivity toward apples relative to candy bars (thus replicating Karpinski & Hilton's [2001] findings). However, we predicted that this positivity would be less apparent on a personalized IAT. Also, although participants might still be motivated to present themselves in a socially desirable light by claiming to engage in healthy eating habits, the inclusion of explicit measures in this less socially sensitive domain allowed us to test the prediction that a personalized IAT would correlate better than a traditional IAT with explicit measures of attitudes, past behavior, and behavioral intentions.

Method

Participants. Sixty-two undergraduates at a Midwestern university participated for course credit. Three were omitted from analyses for committing a large number (> 20%) of errors or because of missing data, resulting in a final sample of 35 women and 24 men.

Materials and procedure. Participants were seated in individual cubicles and read a set of instructions that described the IAT as being about "categorization skills." They were randomly assigned to either the traditional or personalized IAT condition (IAT type was manipulated just as in Experiment 1, which included changes in both normatively pleasant and

unpleasant items and category labels). Parameters of the IAT were modeled after Karpinski and Hilton (2001), with some minor exceptions noted below. Participants were told that they would be categorizing a variety of different items, that instructions on the screen would describe to them how to categorize the items, and to press any key to begin. There were seven blocks in the IAT. The pleasant–unpleasant and liked–disliked items were the same as those used in the first experiment. The apple- and candy bar–related items consisted of words related to the two categories (e.g., *Snickers*, *Red Delicious*). Some of these items were taken from Karpinski and Hilton, and others were derived from our own pretesting (the complete list is presented in the Appendix). The first two blocks consisted of practice with the categorization first of candy bar– and apple-related items, and then pleasant and unpleasant items, respectively. Blocks 3 and 4 were critical combined blocks, where candy bar–related items were associated with the positive category, and apple-related items were associated with the negative (or vice versa, depending on the counterbalancing conditions to which participants were assigned). Block 5 was a practice block consisting of candy bar– and apple-related items. Blocks 6 and 7 were also critical combined blocks, and were identical to Blocks 3 and 4, but the food that was associated with the positive category was now associated with the negative (and the food that was associated with the negative category was now associated with positive).

After completing the IAT, participants completed several explicit measures of their attitudes toward apples and candy bars, which were introduced as "measures of certain beliefs that might affect the categorization skills in which we were interested" (some of which were taken from Karpinski & Hilton, 2001). These included several semantic differential items (*Ugly–Beautiful*, *Bad–Good*, *Unpleasant–Pleasant*, *Foolish–Wise*, and *Awful–Nice*), Liking ("How much do you like eating apples [candy bars]?"), a behavioral measure ("Do you eat apples [candy bars] often?"), and a forced-choice measure of behavioral intention ("If given a choice between an apple and a candy bar, which would you choose?"), all using 7-point scales. Next, they completed a feeling thermometer (on a 0–100 scale) regarding the extent of their favorability toward several filler foods along with our foods of interest. Finally, participants provided rank order information on their preferences for these foods. They were then debriefed, thanked, and dismissed.

Results and Discussion

Effects of IAT version. Critical block means were derived as in Experiments 1 and 2 and were submitted to a 2 (block type: apple/+ vs. apple/−) × 2 (IAT type: traditional vs. personalized) ANOVA, with repeated measures on the first factor. Only a marginal Block Type × IAT Type interaction emerged, $F(1, 57) = 2.90$, $p = .09$. On the traditional IAT, the mean of response latencies on the apple/− task was 834.81 ($SD = 120.30$) compared with 780.23 ($SD = 141.80$) on the apple/+ task. On the personalized IAT, the apple/− task mean was 947.50 ($SD = 200.71$) compared with 950.89 ($SD = 197.28$) on the apple/+ task. Accuracy on the critical apple–candy bar trials did not differ as a function of IAT version (96% for each, $t < 1$). Given that the manipulation included varying whether the attribute items were or were not consensually valenced, accuracy on the attribute trials cannot be compared. However, 48% of the attribute items were categorized positively in the traditional version, with 53% categorized as such in the personalized version.

As in the earlier experiments, attitude estimates for the two versions of the IAT were computed. Reflecting the results of the ANOVA, participants who performed the traditional IAT appeared to prefer apples more than participants who performed the personalized IAT, $t(57) = 1.76$, $p = .08$. As indicated in Table 1, the traditional IAT revealed a significant generalized preference for apples (replicating Karpinski & Hilton, 2001), whereas the personalized IAT suggested no clear preference for one over the other.

Table 1
Descriptive Data and Tests for Each Measure (Experiment 3)

Measure	*M* (*SD*)	*t*(*df*)
Traditional IAT	54.57 (98.89)	2.58(25)*
Personalized IAT	−3.39 (170.80)	< 1(32)
Semantic Differential	0.95 (1.07)	6.83(58)**
Liking	0.37 (1.83)	1.56(58)
Eating Behavior	0.61 (2.25)	2.08(58)*
Behavioral Intention	0.47 (1.89)	1.93(58)
Feeling Thermometer	7.74 (31.03)	1.97(58)
Ranking (reverse scored)	0.81 (3.14)	1.99(58)*

Note. Higher numbers indicate more positive responses to apples relative to candy bars. IAT = Implicit Association Test.
* $p < .05$. ** $p < .01$.

Explicit measures. The semantic differential items were highly related for each attitude object (αs = .72 for apples and .79 for candy bars), so the items were averaged for each attitude object. For the semantic differential and the other explicit attitude measures, difference scores were calculated such that higher numbers indicated a preference for apples over candy bars in order to be comparable to the IAT. These included the Liking, Eating Behavior, and Behavioral Intention questions, the Feeling Thermometer, and the Ranking measure (which was first reverse scored to make it consistent with the other items). Means and *t* tests against 0 are presented in Table 1. Most explicit measures revealed a mild preference for apples over candy bars, although only significantly so for three of the six measures.

Our main goal regarding the explicit measures was to compare IAT-explicit correlations for the traditional and the personalized IAT. Consistent with our predictions, all of the explicit measures correlated significantly with the personalized IAT (see Table 2). However, the explicit measures correlated only weakly and nonsignificantly with the traditional IAT.

The explicit measures all correlated significantly with one another. With the exception of the semantic differential, whose correspondence with the other measures was moderate (*r*s = .44–.62), correlations between explicit measures were quite high (.57–.87). Hence, a single index of explicitly measured attitudes was derived as the average of the standardized individual measures ($\alpha = .91$). As indicated in Table 2, this composite index also correlated strongly with the personalized IAT and only nonsignificantly with the traditional IAT. Tests of the difference between the two IATs' correlations with the explicit measures revealed statistically significant differences for several of the individual measures as well as the overall composite measure. Table 2 provides the results of these tests for each measure.

Recall that in Experiments 1 and 2, IAT scores were also calculated using the algorithm prescribed by Greenwald et al. (2003), except that we did not implement an error penalty, and we used transformed latencies. Experiment 3 was modeled after a number of IAT studies more recent than the original Greenwald et al. (1998) work. Hence, the length of the task had been reduced to 20 practice and 40 critical trials for each of the combined task blocks, just as was true for the Web-based IATs examined by Greenwald et al. Their scoring algorithm dictates that the *D* score

Table 2
Correlations Between Explicit Measures and the Two IATs (Experiment 3)

Explicit measure	Traditional	Personalized	Test of difference: *Z*	Test of difference: *p*
1998 scoring algorithm (difference score from critical blocks)				
Semantic Differential	−.06	.38*	1.63	.10
Liking	.06	.47**	1.60	.11
Eating Behavior	.13	.46**	1.30	.19
Behavioral Intention	.12	.60**	2.03	.04
Feeling Thermometer	.09	.59**	2.09	.04
Ranking (reverse scored)	−.01	.66**	2.85	< .01
Composite	.09	.69**	2.69	< .01
Modified 2003 scoring algorithm (difference scores from practice and critical blocks weighted by pooled standard deviation)				
Semantic Differential	.01	.42*	1.62	.11
Liking	.08	.49**	1.65	.09
Eating Behavior	.15	.43*	1.11	.27
Behavioral Intention	.22	.61**	1.75	.08
Feeling Thermometer	.11	.57**	1.94	.05
Ranking (reverse scored)	.07	.66**	2.61	< .01
Composite	.15	.67**	2.38	.02

Note. IAT = Implicit Association Test.
* $p < .05$. ** $p < .01$.

(mean difference divided by the pooled standard deviation) be computed separately for the practice block and for the critical block and that these two scores then be averaged as the IAT index. However, given the differing lengths of the practice and critical blocks, this practice means that trials on the practice blocks are given twice the weight of those on the critical blocks. We view assigning this disproportionate weight to practice trials to be unwarranted, so in adopting the algorithm to Experiment 3, we computed difference scores and standard deviations such that all trials received equal weight (i.e., means and standard deviations were computed across all 60 trials). Using this algorithm, the traditional IAT ($M = 0.17$, $SD = 0.35$) reflected a greater preference for apples over candy bars than did the personalized IAT ($M = 0.01$, $SD = 0.35$), $t(57) = 1.75$, $p = .09$, just as in the main analyses. Moreover, the pattern of the differences in correlations between the two versions of the IAT and explicit measures remained largely unchanged (see lower panel of Table 2). As before, the correlations were higher for the personalized than for the traditional version.

It is interesting to note that of all the measures, only the personalized IAT failed to reveal at least some preference for apples over candy bars. Given that the explicit measures revealed trends indicative of a preference for apples, this finding may appear problematic to our assumption that the personalized IAT is more likely to tap one's evaluations and less likely to be contaminated by extrapersonal associations compared with the traditional IAT. However, it is likely that participants were still affected by social desirability motives when responding to the explicit measures, which may have been partially responsible for the general preference for apples. It is important to note that this tendency to bias one's responses in favor of apples (in order to appear healthier) on the explicit measures appears separate from the effects of extrapersonal associations on the traditional IAT. If this apparent preference for apples was genuine and the explicit measures and the traditional IAT were both tapping such a preference, then we would expect them to correlate to some extent. That this correlation was not observed suggests that the preference for apples on both types of measures was affected (at least partially) by two different factors—extrapersonal associations in the case of the traditional IAT and social desirability in the case of the explicit measures.[5]

The effect of extrapersonal associations appears to be differential across individuals. That is, some people were apparently more affected by extrapersonal associations than others, which then disrupted the rank ordering of individuals in the preferences derived from the traditional IAT. Had everyone been affected equally, only an overall shift in the distribution of attitude indices would have been found, and any correspondence between the two measures would not have been interrupted. Apparently, respondents varied in the extent to which they were affected by extrapersonal knowledge when solving the mapping problem posed by the IAT.

In sum, our key predictions for Experiment 3 were supported. First, a traditional IAT revealed a greater preference for apples over candy bars. This finding is consistent with our argument that the IAT is contaminated by extrapersonal associations, and it replicates findings from Karpinski and Hilton (2001). Second, our personalized IAT revealed less of a preference for apples—in fact, this version of the IAT suggested that people like apples and candy bars roughly equally. This finding is analogous to the peanut allergy mentioned earlier. In that case, our reasoning suggested that positive extrapersonal associations would lead someone with a peanut allergy to appear relatively more positive toward peanuts on the traditional IAT than on the personalized IAT. In Experiment 3, negative extrapersonal information about candy bars appears to have made participants appear relatively more negative toward candy bars on the traditional IAT than on the personalized IAT. Finally, the personalized IAT correlated more strongly with explicit measures of liking, past eating behavior, and behavioral intentions than did the traditional IAT. These differences in correspondence provide strong evidence that the traditional IAT is contaminated by extrapersonal associations and that the personalized IAT reduces this contamination. In Experiment 4, we replicated and extended these findings to a different attitudinal domain.

Experiment 4

In Experiment 4 our goals were twofold. First, we hoped to extend the finding of increased personalized IAT-explicit measure correspondence to another attitudinal domain. Second, we hoped to demonstrate that modifications to the IAT's category labels and

[5] Our argument that the explicit measures were still affected by social desirability may appear to defeat the purpose of our using the domain of apples and candy bars over race. However, our reasoning for using apples and candy bars was based on the assumption that people would be more willing to report their attitudes toward apples and candy bars than their attitudes toward Blacks and Whites, not that the domains of apples and candy bars are devoid of all social desirability concerns.

the removal of error feedback alone are sufficient to increase correspondence with explicit measures. Thus, the same normatively pleasant and unpleasant attribute items were used in both versions of the IAT instead of using idiosyncratic items for the personalized IAT (as in Experiment 2). We chose political attitudes (and more specifically, attitudes toward George W. Bush and Al Gore) both because of the importance of the domain and because of prior work demonstrating correspondence between a Bush–Gore IAT and explicit measures of these attitudes (e.g., Greenwald et al., 2003). We expected to see greater correspondence with explicit measures with the personalized IAT compared with the traditional IAT. However, we were reluctant to predict any difference between the two versions of the IAT at the level of the main effect, because even though there is certainly a plethora of positive and negative extrapersonal information available about both politicians, it was not clear to us that the aggregated valence of this information would be positive, negative, or neutral for either politician. In this sense, the political domain differs from the racial domain we examined in Experiments 1 and 2, for which the preponderance of negative extrapersonal associations has been documented (e.g., Devine, 1989). It also differs from the comparison of apples and candy bars in Experiment 3—a comparison that involves a more positive portrayal of the benefits of apples over candy bars (Karpinski & Hilton, 2001).

Procedurally, Experiment 4 was quite similar to Experiment 3, in that participants completed either a traditional or a personalized IAT followed by several explicit measures of attitudes toward various politicians. Measures of behavioral intentions and voting behavior during the 2000 presidential election involving Bush and Gore were also administered.

Method

Participants. Individuals were recruited from student newspaper advertisements and flyers posted on the campus of a Midwestern university for participation in this and other unrelated experiments in exchange for $20 during the summer of 2003. Forty-nine individuals, all of whom had listed hometowns within the United States on a preliminary background questionnaire, served as the participants. Of these, data from 1 participant were omitted because of high errors on the IAT (> 20%) and from another because of equipment failure, resulting in 18 male and 29 female participants.

Materials and procedure. Instructions and procedures were analogous to those of Experiment 3, but parameters and stimuli for the IAT were modeled after Greenwald et al. (2003). The pleasant–unpleasant items for both versions of the IAT were normatively pleasant and unpleasant, and the Bush and Gore items consisted of their full names and last names only in black capital letters and two head-shot photos each of Bush and Gore. There were seven blocks in total, with practice blocks consisting of 20 trials each and critical blocks consisting of 40 trials each. Blocks 1 and 2 consisted of practice categorizing Bush and Gore items (Block 1) and pleasant and unpleasant items (Block 2). Block 3 was a practice combined block, where Bush was associated with the positive category and Gore was associated with the negative category (or vice versa, depending on the counterbalancing procedure), and Block 4 was the critical version of this combined block. Block 5 was a practice block consisting of Gore and Bush categorization only. Block 6 was a practice combined block and was identical to Block 3, but the politician that was associated with the positive category was now associated with the negative (and the politician that was associated with the negative category was now associated with positive), and Block 7 was the critical version of this combined block. Participants completed either the traditional version of this IAT, where the labels "Pleasant" and "Unpleasant" were used, or the personal version, where the labels "I like" and "I don't like" were used. Error feedback was also omitted from the personalized IAT.

Participants then completed several explicit measures of their attitudes toward Bush, Gore, and other politicians. First, participants rated both Gore and Bush using several semantic differential items (*Unattractive–Attractive, Bad–Good, Unpleasant–Pleasant, Foolish–Wise,* and *Awful–Nice*). Next, participants responded to the following five direct comparison questions on a 7-point scale anchored by *Bush* and *Gore*: "Who do you think is more intelligent?"; "Who is more qualified to be president?"; "Who do you think is more likeable?"; "Whose character makes him better suited for the presidency?"; and "If an election involving Bush and Gore as candidates for president were held today, for whom would you vote?" Measures of liking of Bush, Gore, and six filler politicians were then administered using a 7-point scale (0 = *Not at all* and 6 = *Very much*). Next, a feeling thermometer that included Bush, Gore, and six filler politicians that was analogous to Experiment 3 was administered. Participants then reported whether they voted in the 2000 presidential election and, if so, for whom they voted. Finally, participants reported their party affiliation on a 7-point scale anchored by *Definitely Republican* and *Definitely Democrat,* with the scale midpoint indicating *Neither/No preference.* They were then debriefed, paid, and dismissed.

Results and Discussion

Effects of IAT version. After dropping the first two trials of each block and log-transforming latencies, Gore/− and Gore/+ block means were submitted to a 2 (block type: Gore/+ vs. Gore/−) × 2 (IAT type: traditional vs. personalized) ANOVA, with repeated measures on the first factor. No effects were revealed. Thus, participants were no faster on either the Gore/− or Gore/+ blocks, this effect did not differ by IAT type, and participants were no slower or faster on the traditional versus the personalized IAT. On the traditional IAT, the mean of response latencies on the Gore/− task was 775.97 (SD = 155.33) compared with 748.54 (SD = 143.18) on the Gore/+ task. On the personalized IAT, the Gore/− task mean was 803.65 (SD = 264.37) compared with 784.93 (SD = 204.09) on the Gore/+ task.

As in the previous experiments, attitude estimates from the two versions of the IAT were computed such that higher numbers indicate a preference for Gore over Bush. Mirroring the ANOVA results, participants did not appear to prefer either politician whether they completed the traditional ($t < 1$) or the personalized ($t < 1$) IAT, and attitude estimates based on the two versions of the IAT did not differ ($t < 1$). Participants were 97% accurate on critical Bush–Gore trials on the traditional version and 95% accurate on critical Bush–Gore trials on the personalized version ($t < 1$). Given that the same attribute stimuli were used for both versions of the IAT, it was also possible to compare accuracy on these trials as well, which did not differ (accuracy was 97% and 96% for the traditional and personalized versions, respectively; $t < 1$).

Explicit measures. Both the Semantic Differential and Comparison items showed strong internal consistency (all αs > .80), so averages were computed for each of the Bush and Gore Semantic Differential items as well as the Comparison questions. Difference scores were computed for all explicit measures of attitudes (with the exception of the direct comparison average) such that positive numbers indicate a preference for Gore over Bush. The direct comparison index was scored such that positive numbers indicate a preference for Gore (and negative numbers a preference for

Bush). For the political affiliation item, a positive number indicates an orientation favoring the Democratic party. Means for these measures and *t* tests against the null value of 0 are presented in Table 3. As can be seen from the table, the present participants appeared to favor Gore, and they also reported a rather liberal political orientation. Unfortunately, only 27 (56%) of the participants actually voted in the 2000 presidential election. Of these, 14 voted for Gore, 7 voted for Bush, and 6 voted for someone else.

Our key prediction was that these explicit measures would correlate more strongly with the personalized IAT than with the traditional IAT. The prediction was confirmed. Although explicit measures of attitudes toward Bush and Gore correlated with the traditional IAT (thus replicating Greenwald et al., 2003), the personalized IAT showed much better correspondence with explicit measures. Tests of differences in correlations between the two versions of the IAT, presented in Table 4, support this conclusion for nearly all of the measures. As in Experiment 3, a single index of explicitly measured attitudes was derived as the average of the standardized individual attitude measures (which included the Semantic Differential, the Comparison items, Liking measure, and Feeling Thermometer). This composite index also correlated significantly more strongly with the personalized IAT than with the traditional IAT (see Table 4). The same was true for the measure of party affiliation. The correlation between the IAT and actual voting behavior among those who voted in the 2000 presidential election for either Bush or Gore was .31 for the traditional IAT ($n = 12$, $p > .3$), and .61 for the personalized IAT ($n = 9$, $p = .08$). Although the correlation was much stronger in the case of the personalized IAT, the relatively small sample sizes provided insufficient power to detect a significant difference between the correlations ($Z < 1$).

IAT scores also were computed according to the modification of Greenwald et al.'s (2003) algorithm implemented in Experiment 3. Just as in the main analyses, the personalized ($M = -0.04$, $SD = 0.50$) and traditional ($M = 0.07$, $SD = 0.43$) versions of the IAT did not differ ($t < 1$), and neither differed from zero (ts < 1). As can be seen in the lower panel of Table 4, the personalized IAT also maintained better correspondence with explicit measures than did the traditional IAT when using this algorithm, albeit somewhat more weakly than what was revealed by the earlier scoring system. Correlations with actual voting behavior, within the small sample of students who voted, also were largely unchanged (rs = .47 and .60 for the traditional and personalized IATs, respectively).

Table 3
Descriptive Data and Tests for Each Measure (Experiment 4)

Measure	*M* (*SD*)	*t*(*df*)
Traditional IAT	27.43 (146.64)	< 1(27)
Personalized IAT	18.73 (181.50)	< 1(18)
Semantic Differential	0.67 (1.66)	2.75(46)**
Direct Comparison	0.54 (1.77)	2.10(46)*
Liking	0.57 (2.36)	1.67(46)
Feeling Thermometer	18.81 (53.89)	2.39(46)*
Party Affiliation	0.49 (1.83)	1.84(46)

Note. Higher numbers indicate more positive responses to Gore relative to Bush on the attitude measures and a stronger affiliation with the Democratic relative to the Republican party. IAT = Implicit Association Test.
* $p < .05$. ** $p < .01$.

Table 4
Correlations Between Explicit Measures and the Two IATs (Experiment 4)

Explicit measure	Traditional	Personalized	Test of difference *Z*	Test of difference *p*
1998 scoring algorithm (difference score from critical blocks)				
Semantic Differential	.37*	.80**	2.20	.03
Direct Comparison	.48**	.80**	1.75	.08
Liking	.42*	.79**	1.97	.05
Feeling Thermometer	.42*	.77**	1.77	.08
Composite Candidate Preference	.45*	.81**	1.96	.05
Party Affiliation	.01	.71**	2.74	<.01
Modified 2003 scoring algorithm (difference scores from practice and critical blocks weighted by pooled standard deviation)				
Semantic Differential	.50**	.77**	1.50	.13
Direct Comparison	.53**	.74**	1.12	.29
Liking	.57**	.78**	1.27	.20
Feeling Thermometer	.56**	.75**	1.07	.28
Composite Candidate Preference	.56**	.78**	1.30	.19
Party Affiliation	.23	.71**	2.63	<.01

Note. IAT = Implicit Association Test.
* $p < .05$. ** $p < .01$.

In sum, our key predictions were again confirmed. The personalized IAT correlated more strongly with explicitly measured attitudes, party affiliation, and voting behavior. It is important to note that this improved correspondence was found even though the same pleasant and unpleasant items were used in both versions of the IAT and even though participants were not slower or less accurate in completing the personalized IAT.

General Discussion

Data from the four experiments reported here suggest that the IAT has the potential to be contaminated by associations that although available in memory are irrelevant to one's evaluation of the attitude object. Moreover, the simple modifications to the attribute category labels and attribute items that we introduced appear to have been successful in reducing the effect of these extrapersonal associations. In Experiments 1 and 2, White participants appeared less prejudiced on the modified Black–White IAT than on the traditional IAT. This finding is consistent with the reasoning that when completing a traditional Black–White IAT, information about society's negative portrayal of Blacks facilitates the process of assigning Blacks and unpleasant items to the same response key, hence creating more prejudiced attitude estimates. The modified IAT, on the other hand, appeared to reduce the influence of these negative extrapersonal associations. These findings were extended in Experiment 3. Here, participants completing a traditional IAT designed to assess evaluations of apples relative to candy bars showed a clear preference for apples, replicating work by Karpinski and Hilton (2001). However, a modified IAT revealed little preference for one over the other, suggesting that people's personal evaluations of apples are not as positive as would be implied by such societal information as "an apple a day

keeps the doctor away." Moreover, a traditional IAT bore no relationship to explicit measures of apples and candy bars, whereas the personalized IAT correlated highly with explicit measures of preference, behavioral intentions, and reports of past behavior. In Experiment 4, this differential IAT-explicit measure correspondence was replicated in the domain of political attitudes, specifically, toward George W. Bush and Al Gore. Although both versions of the IAT correlated with explicit measures, the personalized version correlated significantly more strongly in almost every case, with some correlation coefficients over .80. Moreover, the personalized IAT correlated more strongly with voting behavior from the 2000 presidential election, though smaller sample sizes prevented this effect from reaching significance.

Other findings from the present experiments help to rule out alternative explanations for the differences observed between the two versions of the IAT. Specifically, that participants showed equivalent error rates in the two IATs suggest that these effects were not driven by a reduction in accuracy motivation or a confusion about which items were "category" items and which were "attribute" items. Also, because participants used the "I like" and "I don't like" keys roughly equally in all experiments, it is unlikely that they ignored one of the keys in solving the mapping problem in order to facilitate categorizing the items in the personalized IAT. Moreover, in Experiments 2 and 4, the same pleasant and unpleasant items were used in both versions of the IAT. That we observed a reduction in prejudice (Experiment 2) and an increase in correspondence with explicit measures (Experiment 4) with the personalized IAT, despite the fact that both IATs used the exact same stimulus items, indicates that the items themselves are not responsible for the observed effects. Thus, the combination of changes in the category labels and error feedback is sufficient to personalize the IAT. Finally, although participants performed the personalized IAT more slowly in Experiments 1 and 3, where the attribute items did differ, there was no main effect of IAT type on overall response latencies in Experiments 2 and 4. Thus, it appears that it was the use of idiosyncratic items that slowed participants. For our purposes, however, what is most important is that this slowing cannot explain the results of Experiments 2 and 4. Instead, what the personalized IAT appears to do is reduce the impact of extrapersonal associations.

Throughout this article, we have advanced the argument that personal attitudes may stand in contrast to the valence implied by other information that individuals possess, such as cultural knowledge. We are not claiming that the self can exist in a cultural vacuum or that there is a "bright line" between an individual and his or her culture. We merely propose that through unique occurrences, genetics, or some other form of experience that is different from the norm, the content of one's mind can include information that is not modal for the culture itself. In other words, people are not merely passive receptacles of cultural associations. More specific to the present purposes, we propose that one can possess culturally derived information about an attitude object that does not influence one's attitude, as exemplified by the peanut allergy mentioned earlier. Positive associations to peanuts can certainly be found within the fathoms of the mind of a person with an allergy to peanuts, but they do not necessarily impact the evaluation that is activated in response to seeing peanuts in a dish or on the list of ingredients in a candy bar. Nor must extrapersonal associations come from some monolithic "culture." They may also derive from the knowledge that some specific others have an evaluation of an object that differs from one's own.

Admittedly, some of the arguments advanced in the present article are more directly supported by the data than others, and several questions remain unanswered. Although we have tried to be as precise as possible concerning our definitions of personal and extrapersonal associations, we also have noted the ambiguity inherent in classifying any given piece of information as unrelated to an individual's attitude and, hence, warranting reference to an extrapersonal association. In Experiment 1, in which participants completed both traditional and personalized versions of the IAT, a correlation of .37 was observed. This empirical overlap clearly indicates that the traditional IAT is not devoid of the personal perspective. Instead, it seems to allow for influences of both personal and extrapersonal associations, whereas the personalized version attenuates any influence of the extrapersonal. It is, however, important to note that we are drawing inferences about the operation of extrapersonal associations on the basis of the operational modifications we made to the IAT. To clarify the distinct influences of personal and extrapersonal associations on the IAT (as well as other implicit measures), both types of associations will need to be manipulated experimentally. Our lab is currently pursuing this approach.

One interpretation of our findings that we would not endorse is that the personalized IAT's "I like" and "I don't like" labels created a "demand" that artificially elevated the correlations with the explicit measures. We would argue that self-report measures are essentially trustworthy in domains such as food preferences and presidential candidates. Moreover, the effects were not limited to explicit measures of liking. The personalized version produced higher correlations with various behavioral measures—both past eating behavior and behavioral intention regarding the choice between an apple and a candy bar in Experiment 3 and voting intention and even party affiliation in Experiment 4. It seems doubtful, for example, that participants' recollections of their past behavior or their political party identification would have been affected by a stronger desire to appear consistent with the implications of their IAT performance in the personalized IAT condition than in the traditional IAT condition.

Nevertheless, a number of additional questions remain. For example, there are probably cases in which the personalized and traditional IATs might be expected to show more similar patterns than those reported here—perhaps when the attitude object in question is not as visible in the greater society or when there exists little variability in people's evaluations of the object. With respect to relationships to other implicit measures, we might speculate that the personalized IAT might relate better than the traditional IAT to priming measures, at least under certain circumstances (see Fazio & Olson, 2003; Olson & Fazio, 2003). On the other hand, there may be cases where the traditional IAT relates more strongly than the personalized IAT to a phenomenon of interest. For instance, the traditional IAT may relate better to the ease or fluency of processing novel information about an attitude object or the ability to recall recently acquired information about an object—information that, although inconsistent with the attitude, benefits from congruency with the extrapersonal associations (e.g., Jacoby & Whitehouse, 1989; Read & Rosson, 1982). However, any information-processing effect, judgment, or behavior that is driven

primarily by personal associations should be better predicted by the personalized IAT. In any case, what is most important for any measure of attitudes is the relationship between the measures and actual behavior. This question was addressed at least somewhat in the present research with respect to participant reports of eating apples versus candy bars in Experiment 3 and voting behavior in Experiment 4. Still, we encourage researchers to extend this focus to work that examines the prediction of important social behaviors.

Conclusion

Despite the excitement of pursuing some of the questions mentioned above, future work should also keep in mind questions regarding the very meaning of the implicit measures that have grown so popular in social psychology in recent years. These measures are now used so regularly that posing the question "what are they measuring?" might appear to be a step backward. On the other hand, no scientist would deny that measurement tools need to undergo a rigorous validation process before strong theoretical inferences are drawn from the data they generate. At the risk of appearing overly skeptical, we encourage caution in interpreting the results of research using the traditional version of the IAT. To the extent that the measure is being used in a domain that involves extrapersonal associations, the IAT may not reflect individuals' attitudes as much as is desired. Like any lens, the IAT appears to color its contents. The more personalized version of the IAT that we have examined in the present research focuses the IAT on more personal associations. This more precise focus may provide a stronger basis for interpreting the scores and their meaning.

References

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81,* 789–799.

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

Bessenoff, G. R., & Sherman J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition, 18,* 329–353.

Blair, I. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 359–374). Mahwah, NJ: Erlbaum.

Blair, I., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81,* 828–841.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 760–773.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163–170.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37,* 443–451.

Devine, P. G. (1989). Stereotypes and prejudice: Automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–17.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23,* 316–326.

Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation: Data reanalysis and commentary on Bargh, Chaiken, Govender, and Pratto (1992). *Journal of Personality and Social Psychology, 64,* 753–758.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Fazio, R. H. (2000). Accessible attitudes as tools for object appraisal: Their costs and benefits. In G. R. Maio & J. M. Olson (Eds.), *Why we evaluate: Functions of attitudes* (pp. 1–36). Mahwah, NJ: Erlbaum.

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude–behavior consistency, and the 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., & Olson, M. A. (2003). Implicit measures in social cognition: Their meaning and use. *Annual Review of Psychology, 54,* 297–327.

Fazio, R. H., Powell, M. C., & Herr, P. M. (1983). Toward a process model of the attitude–behavior relation: Accessing one's attitude upon mere observation of the attitude object. *Journal of Personality and Social Psychology, 44,* 723–735.

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

Fiske, A. 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* (Vol. 2, pp. 915–981). Boston: McGraw-Hill.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter? The use of implicit associations toward ethnic groups in person judgments. *Journal of Experimental Social Psychology, 37,* 518–524.

Govan, C. L., & Williams, K. D. (in press). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology.*

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., & Nosek, B. A. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79,* 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

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

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

Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit

partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology, 83,* 367–379.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126–135.

Jordan, C. H., Spencer, S. J., & Zanna, M. P. (2002). I love me . . . I love me not: Implicit self-esteem, explicit self-esteem, and defensiveness. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated Social Perception: The Ninth Ontario Symposium* (pp. 117–145). Mahwah, NJ: Erlbaum.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–778.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly, 24,* 163–204.

Kühnen, U., Schieβl, M., Bauer, N., Paulig, N., Pöhlmann, C., Schmidthals, K. (2001). How robust is the IAT? Measuring and manipulating attitudes of East and West-Germans. *Zeitschrift für Experimentelle Psychologie, 48,* 135–144.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81,* 842–855.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32,* 61–69.

Marsh, K. L., Johnson, B. L., & Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie, 48,* 161–175.

McConnell, A. R., & Liebold, J. M. (2001). Relations between the Implicit Association Test, explicit racial attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology, 37,* 435–442.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence of the effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48,* 107–122.

Mitchell, J. A., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132,* 455–469.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83,* 1029–1050.

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

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = Male, Me = Female, therefore Math ≠ Me. *Journal of Personality and Social Psychology, 83,* 44–59.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413–417.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14,* 636–639.

Olson, M. A., & Fazio, R. H. (in press). Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology.*

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

Plant, E. A. & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology, 37,* 486–501.

Read, S. J., & Rosson, M. B. (1982). Rewriting history: The biasing effects of attitudes on memory. *Social Cognition, 1,* 240–255.

Rothermund, K., & Wentura, D. (2001). Figure–ground asymmetries in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie, 48,* 94–106.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27,* 1164–1178.

Sherman, S. J., Presson, C. C., Chassin, L., Rose, J. S., & Koch, K. (2002). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22,* 13–39.

Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and personality.* New York: Wiley.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude–behavior consistency for stigmatized behavior. *Cognition and Emotion, 15,* 207–230.

Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin, 29,* 170–182.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68,* 36–51.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

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

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815–827.

Appendix

Stimulus Items Used in the Experiments

Normative items	Idiosyncratic items	Apple items	Candy bar items
Pleasant	coffee	sauce	Snickers
	disco	cider	Milky Way
caress	spinach	pie	Kit Kat
freedom	storms	red	Reese's
health	tequila	Red Delicious	Hershey's
love	jogging	Golden Delicious	Baby Ruth
peace	opera	Granny Smith	Butterfinger
cheer	cleaning house	orchard	chocolate
friend	garlic	tree	peanuts
heaven	romance novels	seed	nougat
loyal	motorcycles		
pleasure	country music		
	television		
Unpleasant	airplanes		
	football		
abuse	beer		
crash	caves		
filth	Clinton		
accident	fraternities		
death	Monday		
grief			
poison			
stink			
sickness			
murder			

Received March 6, 2003
Revision received December 8, 2003
Accepted December 10, 2003 ■