
A Threat in the Computer: The Race Implicit Association Test as a Stereotype Threat Experience

Cynthia M. Frantz
Oberlin College

Amy J. C. Cuddy
Princeton University

Molly Burnett
Heidi Ray
Allen Hart
Amherst College

Three experiments test whether the threat of appearing racist leads White participants to perform worse on the race Implicit Association Test (IAT) and whether self-affirmation can protect from this threat. Experiments 1 and 2 suggest that White participants show a stereotype threat effect when completing the race IAT, leading to stronger pro-White scores when the test is believed to be diagnostic of racism. This effect increases for domain-identified (highly motivated to control prejudice) participants (Experiment 2). In Experiment 3, self-affirmation inoculates participants against stereotype threat while taking the race IAT. These findings have methodological implications for use of the race IAT and theoretical implications concerning the malleability of automatic prejudice and the potential interpersonal effects of the fear of appearing racist.

Keywords: *Implicit Association Test; stereotype threat; self-affirmation; implicit racial attitudes*

Racial attitudes, politically charged as they are, have challenged social psychological measurement. People's reluctance to self-report true feelings and beliefs has grown, just as overt bigotry has become unacceptable in virtually all social arenas in the United States. To bypass these social desirability limitations, prejudice researchers developed implicit measures of attitudes designed to find unconscious and uncontrollable preferences (for a review, see Greenwald et al., 2002). These state-of-the-art tools seem to have captured not only the prejudiced atti-

tudes of people who would knowingly misreport them on surveys but they have also captured the surprisingly biased attitudes of people who believe themselves to be unprejudiced. As Devine (2001) wrote in her introduction to a recent *Journal of Personality and Social Psychology* issue dedicated to the topic, "Even those who consciously renounce prejudice have been shown to have implicit or automatic biases that conflict with their non-prejudiced values that may disadvantage the targets of these biases" (p. 757). In a sense, these new measures went beyond where researchers hoped and expected they would go, namely, measuring deliberately concealed prejudice. They also have taken us into a back alley, forcing us to confront the unanticipated negative

Authors' Note: Portions of this research were presented at the annual meeting of the American Psychological Society, New Orleans, Louisiana (June 2002). The authors wish to thank the following people for help with data collection: Erica Carranza, Michel Desert, Michelle Garretson, Jack Gelfand, Ming Meng, Robert Mirabile, Michael Norton, Joseph Simmons, and the students of Psychology 101 at Princeton University. We are grateful to Joshua Aronson, J. Kenworthy Bilz, Geoffrey Cohen, Susan Fiske, Brian Nosek, Matthew Schulkind, Joseph Simmons, and Gregory Walton for their valuable comments on this research. Correspondence concerning this article should be addressed to Cynthia Frantz, Department of Psychology, Oberlin College, Oberlin, OH 44074; e-mail: cindy.frantz@oberlin.edu.

PSPB, Vol. 30 No. 12, December 2004 1611-1624
DOI: 10.1177/0146167204266650

© 2004 by the Society for Personality and Social Psychology, Inc.

racial attitudes of people who might otherwise be considered unprejudiced.

The potential of these implicit measures to expose unwanted and unacknowledged prejudice means they also have the potential to provide feedback that damages a person's self-concept. We propose that some White people feel this threat when they believe their racial attitudes are being tested, which undermines their performance on the test. A similar kind of performance-hindering situational threat has been documented in other domains and has been named stereotype threat (Steele, 1997). It is the apprehension experienced by members of a group who feel they might perform in a manner that is consistent with a negative stereotype, especially when the situation is self-relevant. Stereotype threat diminishes performance on tests of intelligence (Steele & Aronson, 1995), math ability (Spencer, Steele, & Quinn, 1999), affective processing (Leyens, Desert, Croizet, & Darcis, 2000), and even on tests of athletic performance (Stone, Lynch, Sjomeling, & Darley, 1999). Just as stereotype threat undermines performance of people in other valued domains, it may also for White people who are concerned about "failing" a test of racism.

In this article, we investigate whether stereotype threat distorts some participants' responses on the race Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), a widespread measure of implicit racial attitudes. We also explore a potential solution to the stereotype threat problem: If the situation is threatening, perhaps affirming important sources of self-esteem (Steele & Aronson, 1995) can alleviate the threat.

THE IMPLICIT ASSOCIATION TEST

The IAT (Greenwald et al., 1998) measures the relative ease with which people make associations between target categories and evaluations. Because this measure was developed as a way to uncover attitudes that may not be under conscious awareness, it is particularly valuable in investigations of highly charged topics, such as racial attitudes, allowing researchers to circumvent the social desirability issues. Because of its ease of use and the robust results it generates, the measure has garnered a lot of attention and is presently being used for a wide variety of research (e.g., Devos & Banaji, 2002; Rudman, Greenwald, Mellott, & Schwartz, 1999; also see the IAT Web site at <http://www.implicit.harvard.edu>). A recent issue of the *American Psychological Society Observer* compared the revolution currently taking place in social psychology as a result of the IAT to the Copernican Revolution spurred by Galileo's findings in the 16th and 17th centuries (Kester, 2001).

The IAT rests on the assumption that response times will be faster (i.e., shorter) when participants are pairing

highly associated concepts and evaluations and that this faster pairing reflects implicit prejudice. The authors explain that in a society such as the present-day United States, with its clear signs of residual racism, White Americans may have an easier time associating White with pleasant words and Black with unpleasant words (Greenwald et al., 1998). The IAT, administered to participants via computer screen, assesses the association between a target concept and an attitudinal dimension. In the race version of the IAT, the first target concepts introduced are Black versus White. People differentiate between exemplars of the two categories (e.g., faces or names) by pressing a left-hand key for one category and a right-hand key for the other. In the second step, the attitudinal dimension is introduced (pleasant vs. unpleasant) and participants differentiate between exemplars of these attributes (e.g., war, cancer vs. peace, flower). The third step superimposes the two categories, with targets and attributes appearing in randomly alternating trials. The fourth step reverses target concepts (i.e., left to right), and the fifth and final step again superimposes targets and attributes on the same sides as in the fourth step. The term "IAT effect" refers to the discrepancy in the speed with which a person responds to the superimposed pairings. In other words, one might either respond faster to the Black + pleasant/White + unpleasant pairing or to the Black + unpleasant/White + pleasant pairing.

According to Greenwald and colleagues (1998), the IAT is a more powerful tool for measuring implicit attitudes than many of the other unobtrusive scales that have been developed. The chief implicit method previously used was priming (e.g., Dovidio, Evans, & Tyler, 1986; Fazio, Jackson, Dunton, & Williams, 1995), but the IAT method is somewhat easier to administer and is more sensitive, allowing for the use of smaller sample sizes and for the detection of smaller distinctions.

CONTEXTUAL CUES

MODERATE IAT RESULTS

Several lines of research have uncovered cues in the experimental context that can affect IAT performance. For example, Dasgupta and Greenwald (2001) found that exposure to admired Black people (e.g., Denzel Washington) and disliked White people (e.g., Jeffrey Dahmer) attenuates pro-White bias on the IAT, relative to exposure to disliked Black and admired White individuals. Similar effects have been found resulting from exposure to counterstereotypic film clips (Wittenbrink, Judd, & Park, 2001), counterstereotypic mental imagery (Blair, Ma, & Lenton, 2001), diversity training (Rudman, Ashmore, & Gary, 2001), and social influence (Lowery, Hardin, & Sinclair, 2001).

All of these context effects demonstrate that priming counterstereotypic thoughts can attenuate the IAT effect. However, these context effects do not appear to be biasing the test's accuracy. Instead, the IAT seems to be correctly measuring temporary changes in associations. But these demonstrations that the IAT effect is vulnerable to situational cues allow for the possibility that other situational influences, such as stereotype threat, also may affect IAT results. If this is the case, it may leave open the IAT to systematic biases that distort true racial attitudes.

A THREAT IN THE COMPUTER

Stereotype threat researchers have demonstrated that negative stereotypes lead to distracting pressure that interferes with a person's performance in the given situation (Steele, 1997). Existing research points to several conditions that are central to its occurrence. First, a self-relevant negative stereotype must exist and be commonly known (e.g., women are poor mathematicians). Second, the individual must be in a situation where there is a risk of conforming to the negative stereotype. Steele (1997) explains that stereotype threat is a situational pressure "in the air" (p. 617) that undermines performance in the negatively stereotyped domain. Third, individuals are more vulnerable to stereotype threat if they identify with the performance domain to which the negative stereotype pertains. Experimenters manipulate stereotype threat by placing domain-identified participants in situations where their abilities in the negatively stereotyped domains will allegedly be diagnosed through testing.

Steele and Aronson (1995) provided the initial demonstration of stereotype threat, finding that Black students who believed a test measured intelligence performed worse than students who did not believe it measured intelligence. Since this initial demonstration, others have shown how stereotype threat applies to a wide variety of stigmatized groups and domains: women on math exams (e.g., Spencer et al., 1999), elderly people on cognitive tests (Hess, Auman, Colcombe, & Rahhal, 2003), and people from low socioeconomic backgrounds on intelligence measures (Croizet & Claire, 1998). Stereotype threat can occur even when stereotypes are not purposefully made salient (Spencer et al., 1999).

The phenomenon is not limited to historically stigmatized groups. As Steele (1997) noted, everyone is a member of a group that is characterized by some sort of stereotype. Stereotype threat has been invoked even among White men, a group that is traditionally unstigmatized (Aronson et al., 1999; Leyens et al., 2000).

Thus, the experience of stereotype threat requires neither explicit activation of the negative stereotype nor a history of marginalization. These are important points in considering how stereotype threat might apply to the race IAT.

HOW STEREOTYPE THREAT APPLIES TO THE RACE IAT

We believe the race IAT engenders some of the essential conditions of stereotype threat: it concerns a domain (racism) in which the experimental population (White Americans) is negatively stereotyped (racist). To the extent that participants believe the IAT is diagnostic of racism, it presents a risk of conforming to the negative stereotype. Thus, our main research question focused on whether White people taking the IAT might feel threatened about appearing racist. We believe that participants who take the race IAT, who are usually college students, are aware of the existence of the cultural stereotype. Furthermore, participants might often believe (whether they are explicitly told or they deduce as much) that the test is diagnostic of their own racial prejudices. Although White people may not perceive their group as one that is generally stigmatized, the literature reviewed above indicates that situational pressure is sufficient to induce stereotype threat and that internalizing the negative stereotype is not required. Previous research supports the contention that many White people are concerned about appearing racist (Devine & Monteith, 1993; Dunton & Fazio, 1997; Greenwald et al., 1998). Vorauer and colleagues (Vorauer, Hunter, Main, & Roy, 2000; Vorauer, Main, & O'Connell, 1998) suggest that members of majority groups believe that minority group members expect them to be prejudiced and that, in turn, majority group members are concerned about appearing prejudiced in intergroup interactions. Therefore, some White people might experience a sense of threat when they believe their racial attitudes are being evaluated.

At first blush, the IAT may look very different from the kinds of tasks typically used in stereotype threat research. One might argue, for example, that the IAT measures an individual difference (the strength of implicit associations) rather than performance. Prior to stereotype threat research, a similar argument might have been made about intelligence tests and the math SAT (that they measure individual differences in intelligence or math knowledge, rather than performance). Stereotype threat research has taught us that many tasks designed to measure stable individual differences are actually often influenced by the social context. We believe that the IAT is another such measure.

Another way in which the IAT may look different from traditional stereotype threat tasks concerns the level of cognitive processing involved. Stereotype threat research has typically involved tasks that require higher order cognitive processing, and researchers have argued that anxiety over confirming a negative stereotype creates a cognitive load that impairs performance on the domain-relevant task (Croizet et al., in press). The IAT does not involve elaborate cognitive processes. However, stereotype threat also can affect athletic performance (Stone et al., 1999), suggesting that elaborate cognition is not a prerequisite of stereotype threat.

SELF-AFFIRMATION THEORY

The core of the stereotype threat phenomenon is the situation's potential to damage the target's positive self-concept. Perhaps if the self-system can be bolstered in some way, the situation may not have a deleterious effect on performance. Claude Steele also has addressed this issue in his work on self-affirmation (Steele, 1988; Steele, Spencer, & Lynch, 1993). Normally, the term "self-affirmation" is used to describe situations that yield reassuring thoughts of the self within domains irrelevant to the threatened domain (Tesser, 2000). In this case, one would expect people to affirm themselves after they feel the threat of possibly being labeled as prejudiced by thinking about other ways in which they are moral, considerate, or the like. But Tesser (2000) suggests that self-affirmation is most effective when the aspect of the self that has been threatened is bolstered, although this is often more difficult to accomplish than domain-irrelevant self-affirmation.

Self-affirmation also can reduce threat *before* the fact (Cohen, Aronson, & Steele, 2000; Sherman, Nelson, & Steele, 2000). For example, Sherman et al. (2000) found that previously affirmed participants were more willing to accept and benefit from threatening health information than were nonaffirmed participants. Along the same lines, we hypothesize that shoring up one's egalitarian values prior to having one's racial attitudes measured may bolster the self-system and help ameliorate the performance deficits caused by stereotype threat.

THE PRESENT RESEARCH

The present series of experiments examines whether stereotype threat affects IAT performance. We hypothesize that (a) White people are aware of a stereotype that they are racist and (b) when they believe the IAT is a test of racism, their IAT scores will inflate (i.e., higher pro-White bias). We further hypothesize that (c) the magnitude of this inflation is moderated by participants' motivation to control prejudiced responses and (d) diminished by self-affirmation. In a preliminary study,

we demonstrate that White people believe that their group is stereotyped as racist. Experiment 1 provides an initial demonstration of a stereotype threat effect on IAT performance. Experiment 2 replicates that effect and illustrates that even when elaborate attempts are made to mask the purpose of the study, many people suspect that the IAT is measuring racism, thus becoming vulnerable to stereotype threat. Experiment 2 also shows that people who identify strongly with the goal of controlling prejudiced responses (*domain-identified* participants) are especially susceptible to stereotype threat when taking the IAT. Finally, in Experiment 3, we demonstrate the power of self-affirmation to ameliorate the effect of stereotype threat on IAT performance.

PRELIMINARY STUDY OF WHITE METASTEREOTYPES

Although literature cited above suggests that White people do indeed fear appearing racist, we collected data to verify that White people believe others stereotype their group as racist. Vorauer and colleagues (1998) refer to these kinds of beliefs—the beliefs people have about what members of other groups believe about their group—as metastereotypes. As discussed earlier, these beliefs can affect how people feel and behave in interracial interactions (Vorauer et al., 1998, 2000). To assess these metastereotypes, the preliminary study asked White participants to report the stereotypes others had about their group.

Method

A questionnaire was posted online and all White psychology majors at Oberlin College were invited by e-mail to complete the questionnaire anonymously. The response rate was 49%, which resulted in 18 responses (61% female).

Participants were first asked to indicate their race and gender. Then they were asked, "What are the negative perceptions that other groups have about your racial group?" Two hypothesis-blind raters coded these responses. They were instructed to determine whether a participant's response described Whites as racist/bigoted/prejudiced and also to determine whether the response described Whites as dominant/powerful/oppressors. The raters agreed 94% of the time (Cohen's kappa = .94). The first author adjudicated disagreements.

Then, participants were asked to compare both Blacks and Whites to other races on a series of qualities (e.g., "family oriented," "materialistic"). Of primary interest was the item, "Compared to other racial groups, are Whites seen as more racist?" They responded on a 5-point scale ranging from 1 (*not at all*) to 5 (*extremely*).

Results and Discussion

The results overwhelmingly support the hypothesis that White students are aware that they are seen as potential racists. On the free response item, 33% explicitly described a stereotype of Whites as racist/bigoted. The other 67% described Whites as oppressors/dominators.¹

When asked about perceptions of racial groups directly, participants indicated that whereas Whites were seen as more racist than other groups ($M = 4.5$, $SD = .51$), Blacks were seen as less racist than other groups ($M = 2.55$, $SD = .78$). This difference was highly significant, $t(16) = 9.29$, $p < .001$, $d = 3.00$. Taken together, the results suggest that White college students are indeed aware that there is a stereotype of Whites as racist.

EXPERIMENT 1

Experiment 1 manipulated whether people believed the IAT was diagnostic of racial attitudes (*explicit threat* condition) or not (*no threat* condition). We also included a control condition, in which no mention of racial attitudes was made (*no information* condition). We hypothesized that (a) participants who believed the test measured racial attitudes would show a stereotype threat effect—they would have elevated IAT scores, relative to those who believed it was not a measure of racial attitudes, and (b) participants who were not explicitly told what the study was about (the *no information* condition) also would show elevated IAT effects relative to the non-diagnostic condition, given that suspicion also might generate threat.

Method

MATERIALS AND APPARATUS

The IAT was presented on Dell computers with 15-in. monitors using “E-prime” software (Version 1.0; Schneider, Eschman, & Zuccolotto, 2002). This IAT was modeled after the race IAT developed by Greenwald et al. (1998), which used response latencies to assess the relative strength of association between targets and evaluations. Participants classified Black and White faces and pleasant and unpleasant words using two computer keys. The faces were obtained by taking photos of undergraduates who were unambiguously either Black or White. To be consistent with other versions of the IAT, the photos were cropped to show the faces only from the forehead to nose. We used the same pleasant and unpleasant stimulus words used by Greenwald et al. (1998).

PARTICIPANTS

Participants were 47 female and 51 male White undergraduates enrolled in an introductory psychology class at Princeton University. They took the experiment as part of a required laboratory component.

PROCEDURE

Participants completed the experiment in groups of 10 to 15. All instructions were administered via computer screen. Participants in the *explicit threat* condition read the following:

The IAT compares your attitudes toward two different racial groups. It is a measure of racial bias. In this experiment, we are interested in measuring your unconscious racial attitudes toward Blacks and Whites as accurately as possible. Research shows that a high proportion of Whites show a preference for White people. This is a challenging task, but it's necessary for the aim of this study. Please try hard to help us in our analysis of individuals' racial attitudes.

In the explicit *no threat* condition, participants read the following:

The IAT is a measure of knowledge of cultural stereotypes. In this study, we are interested in measuring the extent to which people are aware of cultural stereotypes. Research shows that knowledge of cultural stereotypes is not related to (1) personal belief in cultural stereotypes or (2) inter-racial attitudes and behaviors. This is a challenging task, but it's necessary for the aim of this study. Please try hard to help us in our analysis of people's knowledge of cultural stereotypes.

No information condition participants read, “This is a challenging task, but it's necessary for the aim of this study. Please try hard.”

All participants read the same instructions, presented on the screen, for how the computer program worked. Participants were told to maximize speed and accuracy but that making mistakes was okay. Participants then advanced through the IAT at their own pace, responding to stimulus words by putting them into categories. When they responded correctly, the next item appeared on the screen after a short delay. The word *incorrect* appeared on the screen after incorrect responses were entered.

Results

DATA PREPARATION

Reaction times were processed following Greenwald et al. (1998). Any response times below 300 ms were changed to 300 ms and those greater than 3,000 ms were changed to 3,000 ms. In addition, as the first two trials typically yield unusually long response latencies, they were dropped from the final analyses. Reaction times are often converted to log scores, and we did this, but it made no difference to the results. The scores reported here are raw means, which are more readily interpretable. The IAT effect was calculated by subtracting aver-

TABLE 1: Experiment 1: IAT Effect by Threat Condition

	M	SD
Explicit threat ($n = 34$)	174.62 _a	221.35
No instructions ($n = 31$)	119.16 _{a, b}	176.30
No threat ($n = 33$)	81.24 _b	154.14

NOTE: The values listed are the differences in mean response times between incompatible and compatible trials in milliseconds. Mean response times with different subscripts are significantly different at the .05 level.

age response times for the “compatible” trial blocks (when Black faces and unpleasant words and White faces and pleasant words were paired with the same response key) from the average response times for the “incompatible” trial blocks (the pairing of Black faces with pleasant words and White faces with unpleasant words).

STEREOTYPE THREAT EFFECT

We hypothesized that the *no threat* condition would yield the lowest IAT effect and that both the *explicit threat* and *no information* conditions would differ significantly from it. Because it seemed possible that the *explicit threat* and *no information* conditions also might differ from each other, we did not combine these two groups in our planned comparisons. Inspection of the data revealed that the experimental groups did not have equal variances. Those in the *explicit threat* condition had significantly larger variance ($SD = 221.35$) than those in the *no threat* condition ($SD = 154.14$), $F(1, 65) = 4.19$, $p < .05$. Because the assumption of homogeneity of variance was violated, ANOVAs are not reported. Instead, we conducted planned comparisons (corrected for unequal variance where appropriate).

As predicted (see Table 1), participants in the *explicit threat* condition had a significantly larger IAT effect ($M = 174.62$, $N = 34$) than those in the *no threat* condition ($M = 81.24$, $N = 33$), $t(59) = 2.01$, $p < .05$, $d = .49$, unequal variances assumed. Participants in the *no information* condition also had a larger IAT effect ($M = 119.16$, $SD = 176.30$, $N = 31$) than those in the *no threat* condition, although not significantly, $t(62) = .92$, ns , $d = .23$. The *no information* and *explicit threat* conditions did not differ from each other, $t(63) = 1.11$, ns , $d = .28$.

We also conducted one-sample t tests to determine whether each condition’s IAT effect differed significantly from zero. IAT effects were significant for both the *explicit threat* and *no information* conditions, $t(33) = 4.6$, $p < .001$, $d = .79$, and $t(30) = 3.76$, $p < .001$, $d = .68$, respectively. Although the IAT effect was smaller for the *no threat* condition, it still differed significantly from 0, $t(32) = 3.03$, $p < .01$, $d = .53$. Thus, all three conditions yielded significant pro-White IAT effects.

Discussion

As predicted, participants in Experiment 1 who were explicitly told that the IAT measured racial attitudes exhibited a significantly larger IAT effect than those who were explicitly told that the IAT did not measure racial attitudes. This is consistent with other stereotype threat results: Participants who believe a test is diagnostic of ability in an area in which their group is stereotypically weak (i.e., our *explicit threat* condition) perform more poorly than participants who believe the test is not diagnostic (i.e., our *no threat* condition).

The mean IAT effect for the *no information* condition, where the purpose of the task was ambiguous, fell in between the *explicit threat* and *no threat* groups. This finding implies that some participants may have correctly guessed the true purpose of the measure (thereby feeling threatened) while others did not. In other words, the *no information* condition may contain a mix of people who did and did not feel threatened. It is important to identify how many people guess the true purpose of the study and whether those guessers also produce elevated IAT effects.

The increased variance in the *explicit threat* group suggests that participants respond to this threatening situation in very different ways. In general, the extent to which participants are identified with the domain being measured determines the power of stereotype threat (Spencer et al., 1999; Steele, 1997; Steele & Aronson, 1995). In the case of the race IAT, perhaps participants who are highly identified with controlling prejudiced responses find taking the IAT more threatening than those who are less identified with controlling prejudiced responses, which we tested in Experiments 2 and 3.

EXPERIMENT 2

A second experiment investigated two hypotheses. First, we hypothesized that a strong motivation to control prejudiced responses—or, in stereotype threat terminology, identification with the task-relevant domain—moderates the stereotype threat effect, such that in threatening conditions high domain-identified participants will show greater increases in IAT scores than low domain-identified participants. Previous research provides strong evidence that domain identification plays a critical role in moderating stereotype threat (Aronson et al., 1999; Steele, 1997). For example, Aronson et al. (1999) recruited White male participants who were either highly or moderately identified with math. Participants who were highly math-identified performed significantly worse in the stereotype threat condition, whereas moderately identified students performed somewhat better under conditions of threat.

Experiment 2 also investigated the role of guessing the IAT's true purpose. We hypothesized that (a) many participants, while taking the IAT, suspect that it is a test of racial bias and (b) this suspicion is equally threatening to being explicitly told that the IAT measures racial bias. In this experiment, we created an elaborate cover story and added distractor tasks to simulate efforts by researchers to mask the true purpose of the IAT. Thus, this experiment provides a rigorous test of whether people guess that the IAT is a measure of racial bias.

Methods

PARTICIPANTS

In exchange for course credit, 86 undergraduate students enrolled in introductory psychology at Amherst College were recruited to participate in a study on categorization in which they would be asked to categorize different groups of objects (e.g., faces, cars, shoes, fruits and vegetables) by color. We eliminated data from 22 participants who completed only one part of the two-part study and 13 non-White participants. The final sample included 33 women and 18 men (total $N = 51$).

MATERIALS AND APPARATUS

IAT. The same computer program used in Experiment 1 was used in Experiment 2, with one modification: Unrelated categories were interspersed between the five race blocks. These category discriminations included red and silver cars, green and yellow vegetables, black and white shoes, loud and quiet attributes, as well as combined versions of these tasks. In addition, a second version of the test switched the third and fifth blocks to control for order effects. Only data from the race blocks were analyzed. The program was again administered on Dell computers with 15-in. monitors.

Domain identification. Dunton and Fazio's (1997) scale of Motivation to Control Prejudiced Reactions (MCPR) served as our measure of domain identification. The MCPR includes items such as, "I get angry with myself when I have a thought or feeling that might be considered prejudiced" and "It's never acceptable to express one's prejudices." Participants rated each item on a bipolar scale ranging from -3 (*strongly disagree*) to $+3$ (*strongly agree*). Reliability was acceptable, $\alpha = .69$.

Suspicion about the IAT. Participants completed a demographics questionnaire that also asked them "What do you think this study is about?"

PROCEDURE

Time 1. Participants completed the MCPR scale, along with other unrelated questionnaires, during class.

Time 2. Several weeks later, students were recruited to complete the laboratory part of the study. Students were

not told that the two sessions were related. Participants were seated in front of a computer that delivered all instructions. Participants randomly received one of two explanations for the purpose of the study. Those in the *masked threat* condition were told, "We are interested in measuring the effects of brain lateralization on color perception and how that affects your ability to categorize." Participants in the *explicit threat* condition read, "We are using a measure called the Implicit Association Test to measure the underlying preference you feel for different categories. Today we will be testing several stimulus categories (vegetables, cars, shoes, words) but we are most interested in your response to Black and White faces. It has been shown that most Whites show a strong preference for White faces and we would like to examine this occurrence."

After receiving this explanation, all participants received instructions, identical to those used in Experiment 1, about how the computer program worked. Participants then advanced through the IAT at their own pace. Finally, participants completed the demographics and suspicion about the IAT questionnaire.

Results

Data were prepared for analysis as outlined in Experiment 1. Once again, all data are reported as raw reaction times. The magnitude of each participant's IAT effect was calculated by subtracting their average compatible reaction time (Black + unpleasant/White + pleasant) from their average incompatible reaction time (Black + pleasant/White + unpleasant). Larger scores indicate greater pro-White bias.

SUSPICION ABOUT THE IAT

Participants' answers to the IAT suspicion question were independently coded as to whether the participant mentioned race. The two coders agreed on 94% of responses (Cohen's kappa = .87). The first author adjudicated disagreements. Seventeen of the 27 participants in the *masked threat* condition (63%) guessed that the study was assessing racial attitudes. This condition was then split into two, resulting in three groups overall: *explicit threat* ($n = 24$), *masked threat* (nonguessers, $n = 10$), and *suspected threat* (guessers, $n = 17$). No significant difference occurred between guessers and nonguessers on the MCPR scale, $t(21) = -1.44$, $p = .17$; M guessers = 0.82, $SD = 0.83$ versus M nonguessers = 0.42, $SD = 0.32$.

THREAT EFFECT

We then examined IAT data according to the three divisions described above (see Table 2 for means and standard deviations). As in Experiment 1, the assumption of homogeneity of variance was violated, making ANOVAs inappropriate. Those in the *explicit threat* group

TABLE 2: Experiment 2: IAT Effect by Threat Condition

	M	SD
Explicit threat ($n = 24$)	167.44 _a	200.10
Suspected threat ($n = 17$)	204.28 _a	137.35
Masked threat ($n = 10$)	53.54 _b	91.73

NOTE: The values listed are the differences in mean response times between incompatible and compatible trials in milliseconds. Mean response times with different subscripts are significantly different at the .05 level.

had a significantly larger variance than those in the *masked threat* group, $F(1, 34) = 5.11, p < .05$. We used planned t tests correcting for unequal variance to compare the groups.

As predicted, those in the *masked threat* condition had significantly smaller IAT effects than those in the *suspected threat* condition, $t(25) = -3.01, p = .005, d = 1.25$, and than those who were explicitly threatened, $t(32) = -2.27, p = .03, d = .68$. Participants in the *explicit threat* group did not differ significantly from those in the *suspected threat* condition, $t(39) = .66, p = .52, d = .21$ (see Table 2 for means).

The average IAT effect for each of these three groups was then compared to zero to determine which groups displayed a significant pro-White bias. The IAT effect for those in the *masked threat* condition—the truly non-threatened group—was marginally different from zero, $t(9) = 1.85, p = .10, d = .58$. The IAT effects for those in the *suspected* and *explicit threat* conditions differed significantly from zero, $t(16) = 6.13, p < .001, d = 1.49$, and $t(23) = 4.01, p < .001, d = .84$, respectively.

DOMAIN IDENTIFICATION

Next, we examined the role of domain identification in moderating the IAT effect. We hypothesized that under conditions of threat, there would be a positive relationship between MCPR score and IAT effect, such that those more concerned about appearing racist would show larger IAT effects. Similarly, we hypothesized that there would be no relationship, or perhaps even a negative relationship, between MCPR score and IAT effect under conditions of low threat. To test these hypotheses, we conducted regression analyses, with specific interest in the Threat \times MCPR interaction. Diagnostics revealed that one data point was unduly influencing the regression equation, Cook's $D = 1.14$. This participant was removed from subsequent analyses.

For regression analyses, we dummy-coded those in the *suspected threat* and *explicit threat* conditions (those who knew the purpose of the IAT) as threat = 1 and those in the *masked threat* condition as threat = 0. Because regression equations including interaction terms often result in collinearity problems, the dummy-coded threat

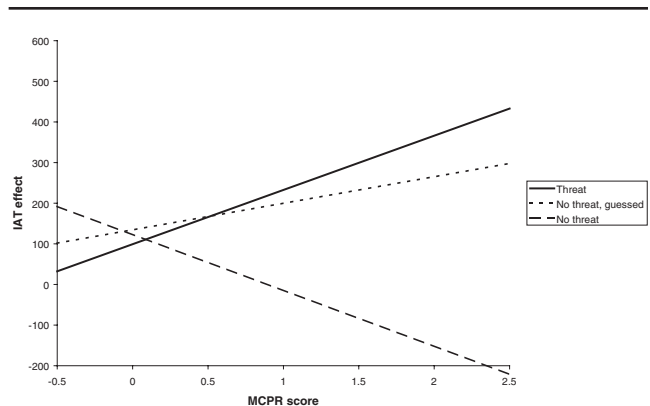


Figure 1 Study 2: The relationship between IAT scores and domain identification as a function of condition.

NOTE: IAT = Implicit Association Test, MCPR = Motivation to Control Prejudiced Reactions Scale.

variable and MCPR were zeroed (by subtracting the mean of the variable from each individual score). The interaction term was computed by multiplying the zeroed variables. The regression included zeroed threat, zeroed MCPR, the interaction, and gender.

Gender was a significant predictor, $b = 115.37, t(38) = 2.08, p < .05$, with women showing higher IAT scores. The Threat \times MCPR interaction was borderline significant, $b = 325.46, t(38) = 1.91, p = .06$. Figure 1 visually depicts the interaction of domain identification and threat condition, with the three conditions shown separately. Our hypotheses were supported; participants in the *explicit threat* and *suspected threat* conditions showed a positive relationship between MCPR and IAT score ($r = .36$), whereas those in the *masked threat* condition showed a negative relationship between MCPR and IAT score ($r = -.52$).

To present the data in a more interpretable form, we calculated the means of those low and high on the MCPR, as determined by a median split. For high MCPR scorers, or high-identified participants, threat status affected IAT scores, such that those who were threatened had a significantly higher mean IAT effect ($M = 239.26$) than those who were not threatened ($M = 4.88$), $t(13) = 4.83, p < .001, d = 1.46$, corrected for unequal variance. For low MCPR scorers, or low-identified participants, threat status did not affect IAT scores, threat $M = 121.00$, no-threat $M = 91.49, t(20) = .35, p > .5, d = .18$.

Discussion

Once again, participants under conditions of explicit threat had elevated IAT scores, supporting the stereotype threat hypothesis. Furthermore, fully 63% of participants receiving a cover story and engaging in multiple distracting categorization tasks still guessed the true pur-

pose of the IAT. The frequency of guessing under these conditions indicates that the purpose of the IAT is easily identified by the majority of participants and suggests that guessing rates may be even higher when it is administered with no attempt to mask its purpose. Because those who did see through the cover story also exhibited a stereotype threat effect, the high rate of guessing is of particular concern for researchers wishing to measure participants' attitudes under allegedly nonthreatening conditions.

An alternative explanation is possible that does not assume a threat effect: perhaps guessers produced large discrepancy scores because the discrepancy itself drew their attention to the IAT's true purpose. If large discrepancies result in guessing, then the masked threat participants would consist primarily of people with small discrepancies, whereas the threatened participants would be a mix of those with large discrepancies and small discrepancies (and on average, have larger discrepancies). Although our data do not directly rule out this possibility, it seems unlikely for two reasons. First, it does not explain the results of Experiments 1 and 3 (see below), in which our nonthreatened conditions were truly randomly assigned. Second, this explanation does not account for the positive relationship between MCPR and IAT scores among threatened participants and the negative MCPR-IAT relationship among nonthreatened nonguessers. To explain these findings one would have to hypothesize a curvilinear relationship between implicit prejudice and MCPR score; this does not seem plausible on theoretical grounds, and it is not supported by the pattern of scores in the explicit threat condition.

Most important, the threat-induced elevation of IAT scores occurs primarily for participants who value controlling prejudiced responses. Thus, participants' scores are differentially biased, depending on their domain identification. If all participants' scores were equally elevated, there would be no methodological problem. However, the measure distorts some scores more than others, inflating the scores of people who are presumably least likely to show a pro-White bias. If the typical administration of the IAT causes the least-biased people to look the most biased, then the predictive utility of the IAT is called into question.

EXPERIMENT 3

Experiment 3 sought to identify a method of administering the IAT that would resist stereotype threat effects. As noted above, self-affirmation prior to a self-esteem-threatening event decreases defensive reactions. In Experiment 3, we once again created a situation in which half of our participants took the IAT under a condition of threat. We also gave half of our participants the oppor-

tunity to affirm their commitment to control prejudiced responses before taking the IAT. We hypothesized an interaction effect: Conditions of threat would impair IAT performance of those who were unaffirmed but not the performance of those who were self-affirmed prior to taking the IAT. We also expected to replicate the moderating effect of domain identification found in Experiment 2.

Method

PARTICIPANTS

One hundred twenty-five undergraduates (57 men, 68 women) from Amherst College participated in the experiment for partial fulfillment of a psychology course requirement or for \$5 compensation. Data from the 25 non-White and two non-American participants were excluded from the final analysis and data from 3 participants were excluded due to computer malfunction. The final data set included 95 participants (45 men, 50 women).

MEASURES

IAT. The same computer program used in Experiments 1 and 2 was used in Experiment 3, without the unrelated categories from Experiment 2. The program was administered on the same Dell computers used in Study 2.

Domain identification. As in Experiment 2, participants completed the MCPR scale.

PROCEDURE

By random assignment, participants completed the MCPR either immediately before (self-affirmed) or immediately after (unaffirmed) taking the IAT. During the computer section of the experiment, all participants read the same explanation for the purpose of the study: "We are interested in measuring the speed and accuracy with which people can make different types of paired associations."

Each participant then received one of two sets of information about the task that they were about to complete. In the *explicit threat* condition, participants read the following:

Among other things, the following task will require you to categorize faces of White and Black people. National samples of American college students produce the following table indicating that the majority of college students show a strong preference to associate White with good and Black with bad.

Strong automatic preference for White faces	66%
Moderate automatic preference for White faces	22%
Little to no automatic preference	7%

Moderate automatic preference for Black faces 4%
 Strong automatic preference for Black faces 1%

Participants in the *no threat* condition read the following:

Among other things, the following task will require you to categorize faces of White and Black people. National samples of American college students produce the following table indicating that the majority of college students show little to no preference to associate White with good and Black with bad.

Strong automatic preference for White faces 4%
 Moderate automatic preference for White faces 14%
 Little to no automatic preference 67%
 Moderate automatic preference for Black faces 12%
 Strong automatic preference for Black faces 3%

After receiving the normative information, which was on the screen until participants pressed the space bar, all participants received identical explanations of how the IAT works (the same as those used in Experiments 1 and 2).

After completing the IAT, participants were asked to recall the statistical data given at the beginning of the IAT (as a manipulation check). Then, participants who had not previously completed the MCPR did so at this time. All participants were then debriefed and dismissed.

Results

The data were processed as described in Experiments 1 and 2.

MANIPULATION CHECK

Participants' responses to the manipulation check were categorized as correct if they placed the largest value in the same category that the normative information had designated as the majority group. Following these standards, 93% of participants were categorized as remembering the data. Three participants in the *explicit threat* condition and 3 participants in the *no threat* condition were categorized as not remembering the data. These participants were removed from analyses. MCPR scores did not differ for those who filled it out before taking the IAT ($M = 74.28, SD = 11.69$) versus after ($M = 73.81, SD = 9.39$), $F(1, 84) = .04, p = .84$.

THREAT EFFECT

Once again, the assumption of homogeneity of variance of IAT scores was violated. The *threatened unaffirmed* condition had significantly larger variance ($SD = 291$) than the other three conditions ($SD = 191$), $F(1, 93) = 5.60, p < .05$. Instead of ANOVAs, a planned contrast, corrected for unequal variance, was conducted. As pre-

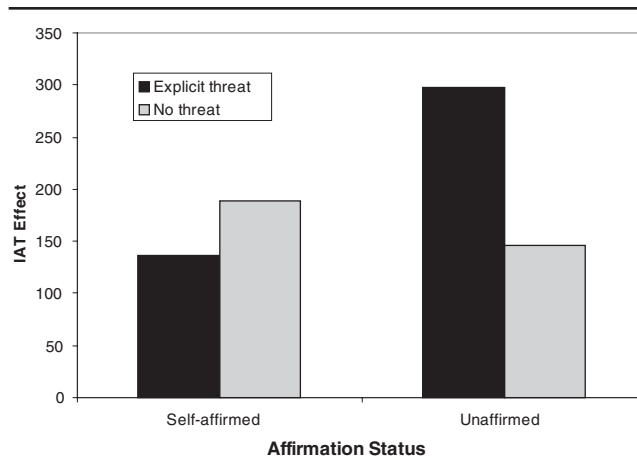


Figure 2 Study 3: IAT effect as response latencies in milliseconds as a function of threat condition and self-affirmation status.

TABLE 3: Experiment 3: IAT Effect by Threat and Affirmation Condition

	N	M	SD
Explicit threat			
Affirmed	23	138.04 _a	180
Unaffirmed	22	298.54 _b	291
No threat			
Affirmed	26	188.89 _a	202
Unaffirmed	24	161.14 _a	195

NOTE: The values listed are the differences in mean response times between incompatible and compatible trials in milliseconds. Mean response times with different subscripts are significantly different at the $p < .05$ level.

dicted, *threatened unaffirmed* participants had a significantly larger IAT effect ($M = 298.54$) than the other three groups ($M = 163.82$), $t(30) = 2.12, p < .05, d = .63$ (see Figure 2 and Table 3 for means and the results of simple pairwise comparisons between conditions). The average IAT effect for each of the four experimental groups was compared to zero. All four showed highly significant IAT effects, t s ranged from 3.69 to 5.02, p s $< .001$.

DOMAIN IDENTIFICATION

Following the procedure used in Experiment 2, regression analysis was used to test whether MCPR scores moderated the threat effect. We hypothesized that those who were threatened and unaffirmed would show a positive relationship between their MCPR and IAT scores, whereas the other three conditions would show either no relationship or potentially a negative relationship (as in Study 2). We created a dummy variable for threat such that the threatened unaffirmed group = 1 and the other

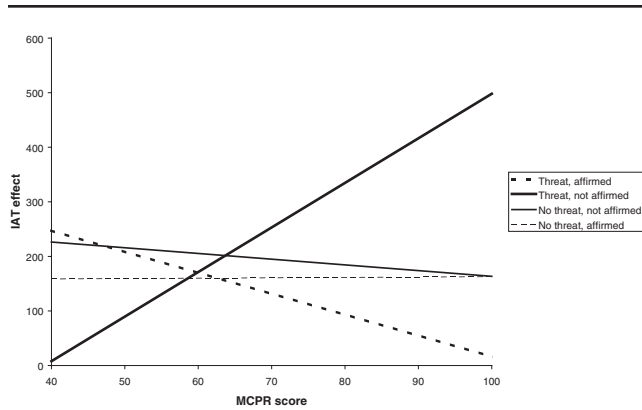


Figure 3 Study 3: The relationship between IAT scores and domain identification as a function of condition.

NOTE: IAT = Implicit Association Test, MCPR = Motivation to Control Prejudiced Reactions Scale.

three conditions = 0. The Threat \times MCPR interaction was the only significant effect, $b = 9.55$, $t = 2.06$, $p < .04$. Figure 3 depicts the relationship between IAT and MCPR scores for all four conditions separately.

For ease of interpretation, we also performed a median split on MCPR score. Among domain-identified (high MCPR) participants, those threatened but unaffirmed had significantly higher IAT scores ($M = 389.44$, $SD = 257.93$) than all other participants ($M = 180.06$, $SD = 221.81$), $t(41) = 2.44$, $p < .05$, $d = .91$. Among low identified participants, IAT scores did not differ between threatened unaffirmed ($M = 200.33$, $SD = 307.96$) and all other participants ($M = 155.03$, $SD = 150.25$), $t(13) = .49$, $p > .5$, $d = .24$, corrected for unequal variance.

Discussion

Study 3 replicates for a third time, with a different manipulation of threat, the central finding of Studies 1 and 2: Participants under conditions of explicit threat, when unaffirmed as in Studies 1 and 2, had elevated IAT scores, relative to those who took the IAT under conditions of no threat or after affirming their egalitarian values. In addition, Experiment 3 also replicated the moderating effect of domain identification, found in Experiment 2: Participants who were highly domain-identified and feeling threatened produced elevated IAT scores, relative to the other groups. Low-domain-identified participants showed no such effect of threat.

In addition, Study 3 demonstrates the effectiveness of domain-relevant self-affirmation for eliminating the stereotype threat effect. Participants who were given the opportunity to assert their egalitarian values before taking the IAT did not show elevated scores, even under conditions of explicit threat. This shows that boosts to

self-esteem, as well as reduction of threat, eliminate the stereotype threat effect.

One significant weakness of Study 3 is that our measure of domain identification (MCPR) also served as our manipulation of affirmation, depending on when participants completed it (before vs. after taking the IAT). Affirmed and unaffirmed participants did not differ on MCPR scores; thus, we have no evidence that participants who took the MCPR after the IAT were motivated to present themselves differently than those who took it before the IAT. This makes using the scores as a moderating variable less problematic. However, if taking the MCPR before the IAT serves as a self-affirmation manipulation, one would expect participants taking it after the IAT to use it as an opportunity to self-affirm and score higher; we did not find this. Ideally, MCPR should be measured in a context separate from the experimental one. Thus, we remain somewhat cautious about the self-affirmation effects. Future research should replicate these findings under more ideal circumstances.

GENERAL DISCUSSION

We hypothesized that (a) White people who are threatened by the possibility of appearing racist will have elevated IAT effects compared to individuals who are not threatened, (b) those who are motivated to control prejudiced responses are more susceptible to this stereotype threat effect, and (c) self-affirmation of one's commitment to control prejudiced responses will attenuate this effect. All three hypotheses were supported; threatened participants (i.e., those who were told or who guessed what the IAT assessed) had higher IAT effects than nonthreatened participants (i.e., those who did not know the nature of the measure).

Furthermore, Experiments 2 and 3 obtained results consistent with previous work on stereotype threat: It is primarily those who identify most strongly with controlling prejudiced responses that had elevated IAT effects in situations that evoke stereotype threat. These results indicate that people who care very much about not appearing racist will likely have increased IAT effects when they think that the study examines racial attitudes.

Finally, Experiment 3 demonstrated that those who were given the opportunity to affirm their values immediately before taking the test did not show elevated IAT effects. Consistent with recent research on self-affirmation, affirming one's values immediately prior to a threatening situation reduces the negative impact of that threat.

Taken together, these findings have important implications for measurement of and theorizing about implicit attitudes.

Methodological Implications

Our data suggest that the race IAT's purpose is extremely difficult to mask and that failing to mask its purpose results in elevated scores. In Experiment 2, participants in the no-threat condition heard an elaborate cover story to conceal the IAT's purpose, yet nearly two-thirds of the participants correctly guessed the aim of the IAT. Moreover, those who guessed showed slightly larger IAT effects than those who were openly told that the experiment measured racial preference.

Ironically, the IAT appears to be the most threatening to people who most want to appear nonracist. In Experiments 2 and 3, unaffirmed participants in the threat condition who had high motivation to control prejudiced reactions had more elevated IAT effects. One would assume that people who believe that acting in an egalitarian manner is important would not have large IAT effects. Indeed, this is the case among those who are truly nonthreatened. But the present results indicate that under situations of threat, the desire to behave in an egalitarian manner leads to larger, not smaller, IAT effects. Thus, under conditions of threat, the IAT does not merely overestimate bias; it *selectively* overestimates bias such that people who are most committed to appearing *unbiased* appear *most* biased.

Consequently, when people take the IAT, what they believe about the nature of the test can affect performance. If individuals are highly motivated to control prejudiced responses, then leading them—or allowing them—to believe that the IAT measures racism will result in elevated IAT effects. Thus, a large IAT effect could imply either high levels of preference for White faces, high levels of concern about appearing racist, or both. The actual meaning of a participant's IAT score is ambiguous.

How can IAT researchers obtain more uniform, interpretable responses? Experiment 1 suggests that stereotype threat can be decreased by telling individuals that the test measures *knowledge* of stereotypes instead of *belief* of such stereotypes. Experiment 3 suggests that informing test-takers that they are unlikely to confirm the negative stereotype alleviates stereotype threat effects. In addition to these methodological adjustments, Experiment 3 also demonstrates the effectiveness of affirming one's egalitarian ideals prior to test-taking. All of these techniques are easy to administer. Other approaches to reducing stereotype threat most certainly exist and should be explored in future research.

Theoretical Implications

The current research contributes to a growing body of research suggesting that implicit stereotyping and prejudice are malleable (for a review, see Blair, 2002).

Researchers have identified numerous moderators of implicit prejudice, including contextual features, participants' goals and motives (e.g., Blair & Banaji, 1996; Dasgupta & Greenwald, 2001; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999), attentional resources (e.g., Macrae, Bodenhausen, Milne, & Calvini, 1999), individual differences (e.g., Lepore & Brown, 1997), and practice or training (e.g., Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). In the research presented here, both contextual features (i.e., threatening or non-threatening instructions) and individual differences (i.e., motivation to control prejudiced responses) moderated implicit biases.

Second, the current research suggests that White people's concerns about appearing nonprejudiced might undermine the quality of their social interactions with members of other racial groups. Many White people are concerned about projecting an egalitarian image (Devine & Monteith, 1993; Dunton & Fazio, 1997; Plant & Devine, 1998). Does stereotype threat lead people to "choke" in interracial social interactions? Researchers have begun to address these issues in studies of interpersonal encounters (Shelton, 2003; Vorauer et al., 1998, 2000), but clearly this is a line of research that calls for continued investigation.

A third theoretical implication of the present findings offers some hope concerning the current state of racism. According to our data, fewer White people are as implicitly biased against Black people as previous IAT data have suggested. The majority of White people who take the test show a significant discrepancy in reaction time between the compatible and incompatible categories; this has been interpreted as a preference for White faces. Although it is likely that many of these people have such preferences, the magnitude of this effect might be falsely inflated for White people who value egalitarianism. Thus, we believe the IAT's findings are less bleak than previous data have suggested.

These results also suggest that individuals are not simply at the mercy of situational factors to determine the impact of threat on their responses. Instead, through the process of self-affirmation, individuals can reduce the threat of a given situation not by changing their definition of the situation (as our experimental manipulations did) but by changing the state of their self-system. To what extent this strategy can be consciously adopted by individuals is a question for future research.

Limitations

These findings, while certainly suggestive, do not allow us to address several important issues. First, because our research exclusively used the IAT, we cannot conclude that the phenomena observed here affect responses on other implicit measures. We would argue that

any situation in which participants identify the measure's (threatening) purpose has the potential to result in stereotype threat effects. However many implicit measures currently in use (e.g., Fazio et al., 1995) do not meet this criterion. Future research should investigate the susceptibility of other implicit measures to these effects.

All three studies produced the same pattern of heterogeneous variances such that variance was substantially higher in the high-threat conditions. We believe this resulted from individual differences that the MCPR scale did not pick up. Extensive research has documented considerable individual differences in how people appraise and subsequently cope with stressful situations and how these differences affect performance (Lazarus & Folkman, 1984). Some of our participants in the high-threat conditions might have appraised the situation as more threatening than others. Other individual differences also might have been at play. For instance, in a recent study of women's math performance, those in the high-threat condition who were particularly self-conscious about their stigmatized status were more susceptible to stereotype threat effects than women who were not stigma conscious (Brown & Pinel, 2003). In short, individual differences that we did not measure likely led to the relatively higher variance in the high-threat conditions across studies.

In addition, this research does not fully explore the conditions under which participants identify the purpose of racial measures, and it does not explore the individual differences that correlate with differences in guessing. We do know from Experiment 2 that relatively elaborate masking procedures do not prevent the majority of participants from guessing the purpose of the IAT. Other experimental procedures also should be tested. In addition, we must recognize the possibility that the participants in Experiment 2 who guessed (our *suspected threat* condition) differ on some unmeasured personality variable that accounts for their elevated IAT scores.

Conclusion

The IAT has become a prominent tool in the investigation of racial attitudes. Because of the IAT's widespread use and methodological strengths, researchers must identify and correct forces that undermine its accuracy. The three experiments reported here demonstrate that stereotype threat can artificially inflate IAT scores (i.e., pro-White bias). Moreover, ironically, people who are motivated not to behave in a prejudiced way are more vulnerable to this type of stereotype threat. These experiments also demonstrate the effectiveness of simple methodological interventions in eliminating stereotype threat. We encourage researchers to focus on using

the race IAT, and perhaps other measures of implicit racial bias, in ways that bar stereotype threat effects from interfering with their accuracy.

NOTE

1. Although neither instructions to students nor to the coders precluded the possibility of a response being coded as mentioning both racism and oppression, students in fact mentioned only one or the other.

REFERENCES

- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.
- Blair, I., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142-1163.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828-841.
- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*, 626-633.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26*, 1151-1164.
- Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.
- Croizet, J. C., Despres, G., Gauzin, M. E., Huguette, P., Leyens, J. P., & Meot, A. (in press). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800-814.
- Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section, *Journal of Personality and Social Psychology, 81*, 757-759.
- Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D. Mackie & D. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 317-344). San Diego, CA: Academic Press.
- Devos, T., & Banaji, M. R. (2002, February). *Do ethnic minorities implicitly view themselves as being less American than Whites?* Poster session presented at the annual meeting for the Society for Personality and Social Psychology, Savannah, GA.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22-37.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudice reactions. *Personality and Social Psychology Bulletin, 23*, 316-326.
- 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.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnum, S. E., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit asso-

- ciation test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 58B, 3-11.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effect of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, 78, 871-888.
- Kester, J. D. (2001). A revolution in social psychology. *The American Psychological Society Observer*, 14, 9-26.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, 72, 275-287.
- Leyens, J. P., Desert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, 26, 1189-1199.
- 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.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Calvini, G. (1999). Seeing more than we can know: Visual attention and category activation. *Journal of Experimental Social Psychology*, 35, 590-602.
- Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, 77, 167-184.
- Plant, E. A., & Devine, P. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811-832.
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81, 842-855.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, 17, 437-465.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes and Intergroup Relations*, 6, 171-185.
- Sherman, D. A. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, 26, 1046-1058.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Steele, C. M. (1988). The psychology of self affirmation: Sustaining the integrity of the self. In R. F. Baumeister (Ed.), *The self in social psychology* (pp. 370-390). Philadelphia: Psychology Press.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, 64, 885-896.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227.
- Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review*, 4, 290-299.
- Vorauer, J., Hunter, A. J., Main, K. J., & Roy, S. (2000). Meta-stereotype activation: Evidence from indirect measures of specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology*, 78, 690-707.
- Vorauer, J., Main, K. J., & O'Connell, G. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, 75, 917-937.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology*, 37, 244-252.

Received July 17, 2003

Revision accepted March 5, 2004