Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?

Russell H. Fazio, Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams
Indiana University Bloomington

The research examines an unobtrusive measure of racial attitudes based on the evaluations that are automatically activated from memory on the presentation of Black versus White faces. Study 1, which concerned the technique's validity, obtained different attitude estimates for Black and White participants and also revealed that the variability among White participants was predictive of other race-related judgments and behavior. Study 2 concerned the lack of correspondence between the unobtrusive estimates and Modern Racism Scale (MRS) scores. The reactivity of the MRS was demonstrated in Study 3. Study 4 observed an interaction between the unobtrusive estimates and an individual difference in motivation to control prejudiced reactions when predicting MRS scores. The theoretical implications of the findings for consideration of automatic and controlled components of racial prejudice are discussed, as is the status of the MRS.

The present research concerns the validity of a technique for measuring racial attitudes unobtrusively. The technique stems from a now widely used procedure, introduced by Fazio, Sanbonmatsu, Powell, and Kardes (1986), for examining the automatic activation of attitudes from memory. The procedure involves priming and permits assessment of the extent to which the presentation of an attitude object automatically activates an associated evaluation from memory. On each trial, the prime that is presented is the name of an attitude object. Its presentation is followed by the display of a positive or negative evaluative adjective. The participant's task is to indicate the connotation of the target word as quickly as possible. Does it mean "good" or "bad"? The latency with which this judgment is made constitutes the dependent measure.

For example, assume that the attitude object *snake* is evaluated negatively by an individual. Presentation of *snake* as the prime may automatically activate the negative evaluation. If the target adjective that is subsequently presented is also negative (e.g., *disgusting*), then the individual is able to indicate the connotation of the target adjective relatively quickly; that is, responding is facilitated. Thus, the technique relies on the presence of facilitation as an indication that the evaluation associated with the primed attitude object has been activated on presentation of the object. Precisely such attitude-congruent facilitation effects have been observed in many experiments using this basic procedure (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, 1993; Fazio et al., 1986; Hermans, De Houwer, & Eelen, 1994; Sanbonmatsu & Fazio, 1986; Sanbonmatsu, Osborne, & Fazio, 1986).

In our first report regarding automatic attitude activation (Fazio et al., 1986), we raised the possibility that the procedure might have utility as an unobtrusive measure of attitude. Essentially, the pattern of facilitation that is exhibited on positive versus negative adjectives can provide an indication of the individual's attitude toward the primed object. Relatively more facilitation on positive adjectives would be indicative of a more positive attitude, and relatively more facilitation on negative adjectives would be indicative of a negative attitude. Furthermore, these estimates are obtained in a situation in which the individual is not aware that his or her attitude is being assessed. During the critical priming task, the participant is not asked to consider his or her attitude toward the object in question. Yet, it is possible to ascertain from the facilitation data the degree to which positive or negative evaluations are activated when the attitude object is presented.

Relatively recent technological advances provide all the more reason to believe that the basic technique might represent a useful and valid unobtrusive measure of attitude. It now is possible to present not words (i.e., not names of attitude objects), but

Russell H. Fazio, Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams, Department of Psychology, Indiana University Bloomington.

This research was supported by Research Scientist Development Award MH00452 and Grant MH38832 from the National Institute of Mental Health and by an American Psychological Association Minority Fellowship. Portions of the data from Studies 1 and 3 were presented at the 1994 and 1995 Midwestern Psychological Association conventions (Jackson & Fazio, 1994, 1995).

We thank Michael Bailey, who developed and tested the software used to present the high-resolution color images as primes and to collect the response latency data. We also thank Teresa Robeson and Jennifer Bradley for so capably serving as experimenters in Studies 1 and 2 and Study 3, respectively, and Edward Hirt for his helpful comments on an earlier version of the article.

Correspondence concerning this article should be addressed to Russell H. Fazio, Department of Psychology, Indiana University, Bloomington, Indiana 47405-1301.

Journal of Personality and Social Psychology, 1995, Vol. 69, No. 6, 1013-1027
Copyright 1995 by the American Psychological Association, Inc. 0022-3514/95/$3.00

high-resolution color images as primes. Thus, a digitized photo of an attitude object—a color photo of a snake, for example—can now be presented as the prime. Using this new technology, our laboratory has typically observed stronger facilitation effects than in our earlier work using words as primes. The present research used such a methodology to investigate racial attitudes.

Racial attitudes provide an excellent context for consideration of the difficulties that can arise when attempting to assess individuals' attitudes. Either because they are unaware of their true sentiments (Banaji & Greenwald, 1994; Greenwald & Banaji, 1995; Nisbett & Wilson, 1977) or because they are reluctant to reveal negativity toward Blacks (e.g., Crosby, Bromley, & Saxe, 1980; Gaertner & Dovidio, 1986; Sigall & Page, 1971), individuals' self-reported attitudes may be suspect. For example, an individual may be reluctant to provide a response that would indicate a negative attitude toward Blacks and, hence, might lead to him or her being labeled as prejudiced. Precisely such concerns about the validity of direct self-reports have motivated pleas for the use of more indirect, unobtrusive measures of racial attitudes (e.g., Crosby et al., 1980; Dovidio & Fazio, 1992).

Various techniques have been promoted over the years as means of circumventing the reactivity of direct measures. The most well known of these is Jones and Sigall's (1971) bogus pipeline technique. By convincing participants that an apparatus, ostensibly recording their physiological responses, is capable of revealing their true attitudes, this technique seeks to induce participants to provide truthful self-reports. Indeed, this technique has been successful in revealing disparities between participants' self-reports on traditional measures and their self-reports when attached to the bogus pipeline (Jones & Sigall, 1971; Sigall & Page, 1971). However, the technique requires an elaborate deception to convince participants of the capacity of the apparatus to discern their true attitudes.

Another type of measure that has been used to examine racial attitudes is the Modern Racism Scale (McConahay, 1986). The Modern Racism Scale was purportedly designed as a nonreactive measure of anti-Black feelings. The scale asks respondents to agree or disagree with a set of beliefs that Whites may or may not hold about Blacks (e.g., "It is easy to understand the anger of Black people in America"). According to McConahay (1986), this instrument allows participants to express negative affect toward a minority without apprehension that this expression might be labeled as prejudiced or racist: "The wording of the items in the Modern Racism Scale permit the expression of negative affect because giving the prejudiced response in each instance can be explained by racially neutral ideology or nonprejudiced race-relevant attributions" (p. 100).

Ultimately, both the bogus pipeline technique and the Modern Racism Scale rely on the individual's self-report. In contrast, the technique we are proposing provides an estimate of participants' attitudes without ever asking them to consider their attitudes. That is, we try to "get inside the head" of the participant. In this respect, the technique represents a potentially bona fide, not a bogus, pipeline.

Various priming paradigms have been used in past research on stereotypes as a means of documenting their automatic activation from memory. Included among these are investigations of racial stereotypes (e.g., Devine, 1989b; Dovidio, Evans, & Tyler, 1986; Dovidio & Gaertner, 1991; Gaertner & McLaughlin, 1983; Gilbert & Hixon, 1991), ageism (e.g., Perdue & Gurtman, 1990), and in-groups versus out-groups (Perdue, Dovidio, Gurtman, & Tyler, 1990). Such research has succeeded in demonstrating the operation of stereotypes at an automatic processing level; stereotype-related constructs were activated by the various primes. In a similar fashion, implicit memory paradigms have recently been used to examine gender stereotyping, demonstrating the occurrence of sexism at an unconscious level (Banaji & Greenwald, 1995). The present research extended beyond such work, however, in that our goal was not merely to demonstrate the activation of differential evaluations as a function of priming but to explore the utility of using each participant's pattern of facilitation as an individual-differences measure. Our initial study aimed to examine the validity of the priming technique as an indirect, unobtrusive measure of attitude.

However, a second, related aim was more theoretical in nature. Devine (1989a, 1989b) has proposed a very influential model of prejudice. Much like our Motivation and Opportunity as Determinants (MODE) model of attitude–behavior processes (Fazio, 1986, 1990), her model focuses on automatic and controlled components of prejudice. The model assumes that high- and low-prejudiced individuals are equally knowledgeable of cultural stereotypes and that these socially shared, cultural stereotypes are likely to be accessible and automatically activated in the presence of a minority group member. What distinguishes prejudiced and unprejudiced individuals is not this automatic component but the controlled component. Nonprejudiced people are presumed to hold personal beliefs that motivate them to inhibit the influences of the automatically activated cultural stereotype. Thus, the model suggests that personal beliefs must be consciously attended to via controlled processes to exert any influence. In contrast, our approach emphasizes the automatic component far more and suggests that there will be meaningful variability in the nature of the evaluations that are activated from memory automatically. We argue that what is automatically activated from memory is not necessarily some socially shared, cultural stereotype but personal evaluations.

In one study testing her model, Devine (1989b) examined the effects of subliminally priming stereotypes commonly held about Blacks in Western culture. The primes included words that were social category labels (e.g., *Blacks* and *Negroes*) and stereotypic descriptors (e.g., *poor* and *athletic*). Participants' perceptions of the ambiguous behavior of a target person whose race was unspecified served as the dependent variable. Devine found that priming of the stereotype associated with Blacks affected participants' perceptions of the target person's behavior. Those participants primed with the cultural stereotype of Blacks rated the ambiguous behavior of the target person as more hostile. This effect occurred equally for both high- and low-prejudiced participants, as identified by the Modern Racism Scale.

These findings raise two questions. First, which concept, the cultural stereotype or one's personal evaluation, is automatically activated in the presence of a minority group member? Devine (1989b) suggested that this commonly held stereotype,

rather than one's personal evaluation, is activated automatically. Yet, her study primed the cultural stereotype of Blacks directly. The primes included words such as *lazy, ghetto,* and *welfare* that tend to evoke negative images. In contrast, priming with the faces of Black individuals would provide an opportunity to examine whether personal evaluations, as opposed to shared cultural stereotypes, are activated on observation of a minority person.

The second issue raised by Devine's (1989b) study concerns the validity of the Modern Racism Scale. Devine found no difference between high- and low-prejudiced participants. Although this lack of a difference may reflect the activation of a similar, culturally shared stereotype (as Devine suggested), it also is possible that the null finding stems from a failure on the part of the Modern Racism Scale to provide a valid measure of an individual's level of racism.

The present study was designed to examine both of these issues, as well as the validity of the indirect technique as a measure of racial attitudes. The automatic activation of attitudes was tested among groups of participants scoring high and low on the Modern Racism Scale. Participants were presented with photographs of faces of White and Black male and female undergraduates and asked to make judgments about the connotation of target adjectives that followed the presentation of each face. If the face (prime) automatically activated evaluations from memory, responding in the adjective connotation task should be affected differently for positive versus negative adjectives. Inferences regarding the participant's attitude toward the individuals represented in the photographs can be drawn from the pattern of facilitation. Relatively greater facilitation on negative than positive adjectives when those adjectives are preceded by Black faces than when they are preceded by White faces would be indicative of a more negative attitude toward Blacks.

More specifically, the goals of the present research were threefold. First, the study explored Devine's assumption about cultural stereotypes versus personal evaluations. In the present study, group membership (i.e., White or Black), rather than the cultural stereotype, was primed. If, as Devine suggested, the shared cultural stereotype is activated in the presence of a minority group, one would expect little meaningful variation in the pattern of facilitation across participants. On the other hand, if it is one's personal evaluation that is activated in the presence of a minority group member, the variation across participants would be more substantial and predictive of race-relevant behaviors.

Such predictive validity constituted the second aim of the present study. The ability of both the Modern Racism Scale and the attitude estimate based on facilitation scores to predict judgment and behavior was examined. To do so, we obtained two additional measures of participants' behaviors and judgments: (a) ratings of an interaction between each participant and a Black experimenter and (b) participants' responses to a number of questions regarding the Rodney King verdict and the ensuing 1992 Los Angeles riots. The data from these two measures were correlated with Modern Racism Scale scores and the unobtrusive attitude estimates to examine the link between the participants' attitudes, as measured by both techniques, and their judgments and behaviors.

Finally, the serendipitous inclusion of a sample of Black students in the study permitted the comparison of patterns of facilitation in the priming task for Black versus White individuals. Thus, estimates of racial attitudes based on facilitation scores could be examined for the degree to which they distinguished Black and White participants. Such discriminability would provide further evidence regarding the validity of the unobtrusive technique.

Study 1

Method

Participants. Fifty-three individuals (45 Whites and 8 Blacks) participated for payment or as partial fulfillment of an introductory psychology course requirement. These individuals were among 479 students who had participated in an earlier mass survey that included the Modern Racism Scale. Scores on the Modern Racism Scale can range from −14 (low prejudice) to 14 (high prejudice). Twenty-five students had scores that placed them among the top 10% of respondents (scores of 2 to 12); 28 had scores that placed them among the bottom 10% (scores of −14). Seven of the 8 Black students were from this latter group.

Stimulus materials. Forty-eight color photographs of White, Black, and other ("other" being Asian and Hispanic) male and female undergraduates served as primes. All photographs were head shots taken against a common background. Volunteers, who were paid $5, signed consent agreement forms permitting the use of their photographs for research purposes. The photographs were digitized as 256 color, 640 × 480 resolution image files. Thirty-two black-and-white yearbook photographs were selected from the 1991 Indiana University yearbook for use in the second and third phases of the experiment described subsequently.

Procedure. On arrival, students were greeted by an Asian female experimenter unaware of their Modern Racism Scale scores. Students were told that the experiment involved word meaning as an automatic skill and that a variety of different tasks would be performed during the experiment. The experimental procedure consisted of six phases, the fourth phase involving the actual priming task.

The first word-meaning task, whose purpose was to obtain baseline data, involved the presentation of a single word on the computer screen. The student's task was to press a key labeled *good* or a key labeled *bad* as quickly as possible to indicate his or her judgment of the word. The list of words consisted of 12 adjectives that were positive in connotation (e.g., attractive, likable, and wonderful) and 12 adjectives that were negative in connotation (e.g., annoying, disgusting, and offensive). Students were instructed to maximize the speed and accuracy of their responses. The order in which the adjectives were presented was randomized for each student. A row of asterisks preceded the presentation of each adjective, serving as a warning signal that the target adjective was about to appear. A given adjective remained on the screen until the student responded or for a maximum of 1.75 s. A 2.5-s interval separated each trial. The student's response was recorded, along with the latency of response (from adjective onset to response) to the nearest millisecond. Students' performance of this task was preceded by a block of practice trials involving different adjectives to familiarize students with the procedure. Students performed two blocks of trials, each block consisting of 24 adjectives. The average latency for the two trials involving any given adjective served as the student's baseline latency for that adjective.

The next two phases of the experiment were intended to prepare students for the priming task that would involve presentation of faces as primes and adjectives as targets. The second and third phases were presented to students as face learning and detection tasks.

The second phase ostensibly involved the ability to learn faces for a later recognition task. The students' task was simply to attend to the faces (targets) presented on the computer screen. They were told that

they would be asked to recall the targets in the next task. The stimuli consisted of 16 black-and-white yearbook photographs of White, Black, and Asian male and female faces. Each photograph was presented twice, once in each of two blocks.

The third phase involved a recognition test of the faces presented in the previous task. The photographs from the previous task were presented on the computer screen. Students were told that their task was to press the key labeled *yes* if the face had appeared in the previous task or to press the key labeled *no* if the face had not appeared in the previous task. Each face remained on the screen for a maximum of 5 s. A 2.5-s interval separated each trial. Students made such judgments about 32 faces, 16 "target" faces that had appeared in the previous task and 16 "filler" faces not previously presented.

The fourth phase involved the actual priming task. Students were told that the previous tasks would now be combined. They were told that our interest was in determining the degree to which the judgment of word meaning was an automatic skill. The experimenter said that if such a judgment was truly an automatic skill, individuals should be able to perform just as well as in the very first phase of the experiment even if they had to do something else at the same time. Thus, students were led to believe that this phase of the experiment involved both the learning of the faces and the judgment of the connotation of the adjectives.

The instructions and procedures were identical to the first task, with one exception. Students were told it was important that they attend to the faces presented because they would be asked to recall the faces in the next task. The row of asterisks was replaced by 48 color photographs of White, Black, and "other" male and female faces. These photographs served as primes. On any given trial, a prime was presented for 315 ms, followed by a 135-ms interval before onset of the target adjective. Thus, the interval between prime onset and target onset, stimulus onset asynchrony, was 450 ms.[1] A 2.5-s interval separated each trial. After an initial practice block involving different faces and adjectives, four blocks of trials were presented. Each block consisted of 48 trials in which each of the primes appeared once, followed by one of the 24 adjectives. Over the course of the four blocks, each prime was paired with 2 positive and 2 negative adjectives. Each Black face and each "other" face were randomly paired with a same-sex White face. The paired faces were followed by an identical set of 4 adjectives. Trials involving the 12 matched pairs of Black and White faces constituted the actual trials of the experiment. The trials involving the 12 "other" faces and their 12 matched White faces served as fillers. These trials were included to reduce the overall proportion of Black faces to which the student was exposed and, thus, minimize the likelihood that students would become aware of the interest in the race shown in the photograph.

The fifth phase of the experiment was the detection task that students had been led to expect during the instructions for the fourth phase. This detection task involved the presentation of the 48 color photographs used during the priming task, along with 48 filler photographs not previously presented. Students were instructed to press the key labeled *yes* if the face was presented in the priming task or the key labeled *no* if the face had not been presented in the priming task. The instructions for this task were identical to those given in the third phase of the experiment. Each photograph appeared on the screen for 5 s or until the student pressed a key. A 2.5-s interval separated each trial. Each photograph was presented once.

The sixth and final phase involved ratings of attractiveness of the color photographs (primes). This phase was intended to bolster the cover story and to provide the basis for the "debriefing" that students were to receive later (see later discussion). Students were told that we were interested in assessing the extent to which the attractiveness of a face determined the degree to which it had distracted them from their task of judging word meaning during the fourth phase. Hence, we were asking them to rate the attractiveness of each photo. Students were instructed to press one of nine keys (1 = *not at all attractive,* 9 = *very attractive*) to indicate their rating of the attractiveness of the photograph. Each photograph appeared on the screen for 15 s or until the student pressed a key.

Additional measures. Two additional measures were collected to examine the possibility that the direct measure (Modern Racism Scale) or an indirect measure (one based on facilitation scores) of students' racial attitudes (or both measures) would be predictive of their behavior. The first measure involved the students interacting with a second experimenter, a Black woman, who was unaware of students' Modern Racism Scale scores. After completing the computer tasks, the students were introduced to the second experimenter with the explanation that this person would take a few minutes to explain the experiment and pay them. During the debriefing session, this experimenter provided an explanation of the experiment, focusing on the possibility of attractiveness as a distraction in the performance of the word-meaning task. The experimenter answered any questions the students may have had at this time. Each debriefing session lasted approximately 10 min. On the basis of this interaction, the experimenter rated each student in terms of friendliness and interest in psychology on scales ranging from *not at all* (−3) to *very* (3). In making these judgments, the experimenter was especially attentive to such factors as smiling, eye contact, spatial distance, and body language. The purpose of these measures was to obtain some indication of the nature of the student's behavior during an interaction with a Black target person and to assess whether a student's Modern Racism Score, indirect attitude score, or both scores would predict such behavior.

As students were leaving the laboratory after the debriefing, the first experimenter stopped each one to ask him or her to complete a survey ostensibly unrelated to this study. This constituted the second measure. Students were asked to complete a questionnaire for a colleague at another university. This "college student national opinion survey" contained a number of items related to the 1992 Rodney King trial and the ensuing Los Angeles riots.[2] The survey asked respondents to indicate the extent to which they agreed or disagreed with statements regarding the trial verdict ("Was it just?") and the anger of the Black community after the verdict ("Was the community justified?"). In addition, respondents were asked to indicate the degree to which they would attribute responsibility for the riots to the following groups: Whites, Blacks, Koreans, the media, and the Los Angeles Police Department. After they had completed the questionnaire, students placed it in an envelope, sealed the envelope, and then placed the sealed envelope in a large manila envelope posted on a bulletin board. This envelope was clearly addressed to a scientist at another university. The students were under the impression that the packet would be mailed to the individual.

Results

Detection data. During the priming phase, students had been instructed to attend to the photographs for recall in a later task. In compliance with the cover story, students later participated in a detection task in which they had to indicate whether

[1] This stimulus onset asynchrony of 450 ms was selected on the basis of earlier pilot testing and experimentation involving photos of objects as primes. Just as in Experiments 2 and 3 of Fazio et al. (1986), the Valence of the Attitude Object × Valence of the Adjective interaction was not observed at a longer stimulus onset asynchrony, but was apparent at the stimulus onset asynchrony of 450 ms. With even shorter prime durations, participants had difficulty identifying and naming the pictures that were presented.

[2] In 1992, four White police officers were acquitted in the brutal beating of a Black motorist, Rodney King. The verdict resulted in looting and rioting in Los Angeles.

a given photograph had or had not been presented during the priming phase of the experiment. These data were examined to check whether the students had followed the instruction to attend to the faces. Detection scores were calculated for each student by subtracting the proportion of false alarms (filler not previously presented that the student incorrectly identified as having been presented) from the proportion of correctly identified hits (previously presented faces correctly identified as such). Performance at chance levels would be indicated by a score of zero on this detection index. Students performed at better than chance levels; their mean score was .65, which was statistically reliable, $t(53) = 26.18, p < .001$.

Facilitation scores. For each student, baseline latency for each adjective was computed from the average of the two presentations of the adjective during the initial asterisk task. The latency for any given target adjective when preceded by a given face was subtracted from the baseline for that adjective to arrive at a facilitation score. Average facilitation scores on positive target adjectives and negative target adjectives were computed for each face. For each student, mean positive and mean negative facilitation scores were computed within each cell of the 2 (race of face) × 2 (sex of face) design. The latencies for any trials on which the student made an error, which averaged 3.52%, were not included in these computations.

Effects of race of student. Of the 53 students, 8 were Black. Thus, an initial analysis examined the effects of the student's race. The data were analyzed in a mixed analysis of variance (ANOVA) involving one between-subjects variable (race of student) and three within-subject variables (race and sex of the photograph and valence of the adjective). This overall analysis revealed main effects for race of the photo, $F(1, 51) = 5.67, p = .021$, and valence of the adjective, $F(1, 51) = 5.29, p = .026$. However, both main effects were qualified by a significant three-way Race of Student × Race of Photo × Valence of Adjective interaction, $F(1, 51) = 25.88, p < .001$. No significant effects concerning sex of the face were observed; therefore, subsequent analyses have collapsed across this variable.

As a means of exploring the nature of the interaction, separate ANOVAs were conducted among the White and among the Black students. The analysis among White students revealed a significant interaction between the race of the prime and the valence of the target adjective, $F(1, 43) = 32.49, p < .001$. The means for this interaction are displayed in the left panel of Figure 1. Greater facilitation[3] occurred when positive adjectives had been preceded by White than by Black primes and when negative adjectives had been preceded by Black than by White primes. This pattern of facilitation suggests that White students, on average, held negative attitudes toward Blacks.

Among Black students, the analysis again revealed a significant interaction between the race of the prime and the valence of the target adjective, $F(1, 7) = 9.36, p = .018$. However, in this case, the pattern of facilitation was in the opposite direction (see the right panel of Figure 1). Greater facilitation occurred when positive adjectives had been preceded by Black than by White primes and when negative adjectives had been preceded by White than by Black primes.

As noted earlier, the differential patterns of facilitation for Black and White students produced a highly significant Race of Student × Race of Photo × Valence of Adjective interaction. This outcome provides the first indication that the unobtrusive technique has some validity. Black and White students displayed very different patterns of facilitation.

Effects of level of modern racism. Subsequent analyses focused on only the White students to allow examination of the effects of level of racism, as measured by the Modern Racism Scale. A mixed ANOVA involving three within-subject variables (race of face, sex of face, and valence of adjective) and one between-subjects variable (level of racism, as defined by the Modern Racism Scale) was conducted. This analysis revealed that level of racism did not interact with any other variable. The critical Race of Face × Valence of Adjective interaction that had been observed earlier was not qualified by level of racism, $F < 1$. For both low- and high-prejudiced White students, greater facilitation occurred when positive adjectives had been preceded by White primes and when negative adjectives had been preceded by Black primes. Thus, the pattern of facilitation held constant regardless of the level of racism, as measured by the Modern Racism Scale. This null effect parallels Devine's (1989b) finding. As noted earlier, the priming effects observed in her study were not moderated by scores on the Modern Racism Scale.

Attitude estimates based on facilitation scores. The facilitation data can be represented in another form, one that proves useful for other purposes. We wished to reduce the facilitation data for any given student to a single index that would serve as an estimate of the student's attitude toward Blacks. Given that multiple observations were available for each student (i.e., facilitation data for 12 White faces and 12 Black faces), it was possible to examine the Race of Photo × Valence of Adjective interaction for each and every student. We computed the effect size of this interaction for each student as our estimate of the student's attitude. In more precise terms, the steps (see Rosenthal, 1991) involved (a) computation of the difference between the average facilitation score for positive adjectives and that for negative adjectives for each and every face, (b) computation of a pairwise t test comparing the difference scores for the White faces and their matched Black counterparts, (c) transformation of the t value to a correlation coefficient, and (d) further transformation of this coefficient via Fisher's r-to-z transformation. For the resulting index, more negative scores reflected a pattern of facilitation indicating greater negativity toward Blacks.

A frequency distribution of these attitude estimates is displayed in Figure 2. Note, first of all, the relative position of the Black versus White students in this distribution. Reiterating what was observed in the ANOVAs, the White students were

[3] Throughout this article, the term *facilitation* is used in a relative sense. To speak of absolute facilitation requires the assessment of a true no-prime baseline that is identical to the primes in such respects as ease of encoding and memory demands (for further discussion of this issue, see Fazio et al., 1986; Jonides & Mack, 1984). In the present case, the task instructions clearly demanded that the students attend to and familiarize themselves with the faces that appeared as primes. Hence, it is not surprising that responding to the adjectives generally took longer during the priming phase of the experiment than during the baseline assessment phase, when the students' sole task was to indicate the connotation of the adjective.

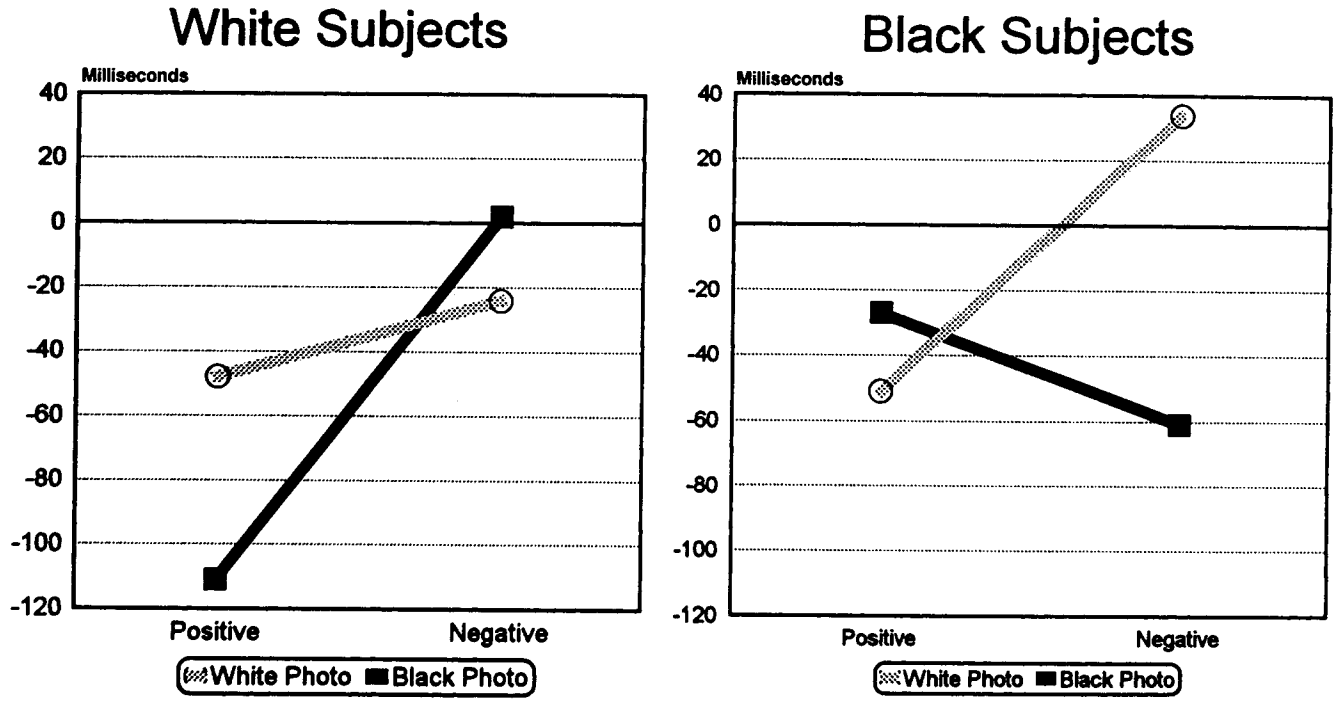


Figure 1. Mean facilitation scores in Study 1 for positive and negative adjectives preceded by photos of White and Black faces.

characterized, on average, by negative attitude scores; their mean score was −.26, which was significantly different from zero, $t(44) = 6.33$, $p < .001$. The Black students were characterized by positive attitude scores; their mean score was .33, $t(7) = 2.51$, $p < .05$. The two means differed significantly from one another, $t(51) = 5.32$, $p < .001$. However, it is also worth noting that there existed considerable variability among the White students. Some of them were characterized by relatively extreme negative attitude estimates. However, some of them were not, and, in fact, some displayed data similar to those displayed by some of the Black students.

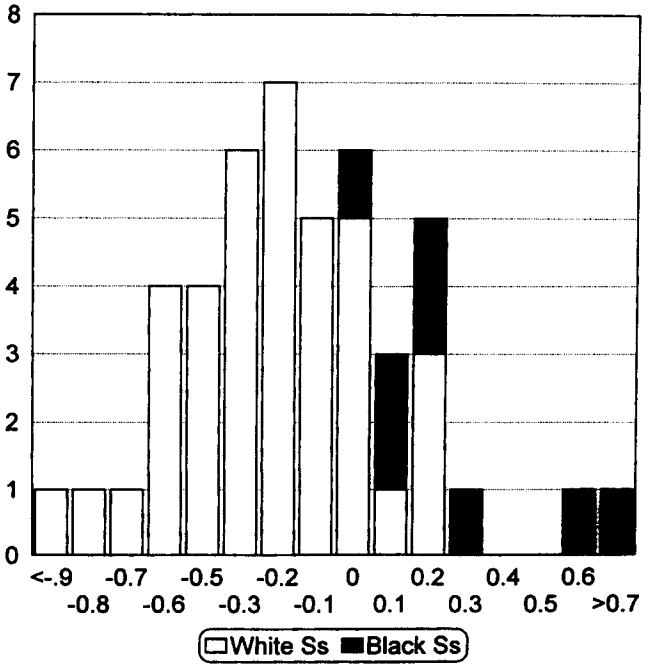


Figure 2. Frequency distribution of attitude estimates among White and Black subjects. More negative scores reflect more negativity toward Blacks.

Predicting judgment and behavior. One of our goals was to examine how well the attitude estimates, as well as the Modern Racism Scale, served as predictors of judgment and behavior. The value of having reduced the facilitation data to a single attitude estimate for each student is that we could easily examine how the unobtrusive estimates correlated with some of the other measures that were collected. The correlations presented were calculated only among the White students. The question of interest is the extent to which the variability that was observed among the White students is meaningful in the sense that it is predictive of other judgments and behavior.

The Black experimenter had rated the friendliness and interest that each student had exhibited during the debriefing. These two ratings, which correlated significantly ($r = .37$, $p < .02$), were standardized, and their average was computed as an indication of the overall tone of the interaction. As shown in Table 1, the indirect attitude estimates correlated significantly with this measure, whereas the level of modern racism (high vs. low, as indicated by whether the student had been selected on the basis of his or her falling in the top or bottom 10% of the distribution of Modern Racism Scale scores in the mass survey) did not.

In addition to the behavioral measure obtained during the debriefing, students' assessments of the Rodney King trial verdict and the subsequent riots were collected in the opinion survey. Students rated how just the verdict was and how justified the anger of the Black community was. The average of these two ratings, which

Table 1
Correlation Matrix of Variables in Study 1

Variable	1	2	3	4	5	6
1. Unobtrusive estimate	—					
2. Modern racism level	.15	—				
3. Interaction rating	.31**	−.09	—			
4. Rodney King verdict	−.06	−.53***	.02	—		
5. Responsibility	.32*	−.41***	.26*	.52***	—	
6. Attractiveness	.18	−.35**	.07	.39***	.20	—

Note. Higher scores on the unobtrusive estimate, the interaction rating, the verdict, and the responsibility and attractiveness measures reflect a more favorable response to Blacks. Higher scores on the modern racism level variable reflect a more negative response.
* $p < .10$. ** $p < .05$. *** $p < .01$.

correlated significantly ($r = .42$, $p = .005$), was computed. This assessment of the verdict correlated well with the level of modern racism but did not correlate with the unobtrusive estimates of attitude based on facilitation scores (see Table 1).

However, the unobtrusive attitude estimates did correlate with students' assignment of responsibility for the escalation of tension and violence that ensued after the verdict. The responsibility measure was simply the difference between the responsibility assigned to Whites and that assigned to Blacks. This measure correlated well with both our unobtrusive estimates and the level of modern racism.

Although attractiveness ratings were collected in the final phase of the experiment largely for the purpose of setting the stage for the Black experimenter's debriefing of the student, the relation of these data to the other variables also was examined. For each student, a point-biserial correlation was computed between a dichotomous variable representing Black versus White photos and the attractiveness ratings. These within-subject correlations were transformed via Fisher's r-to-z transformation. For the resulting index, more negative scores reflected having rated the Blacks as less attractive than the Whites. This measure correlated significantly with scores on the Modern Racism Scale (more prejudiced students were more likely to rate the Black faces as relatively less attractive) but not with our unobtrusive measure (see Table 1).

A factor analysis of these six variables was conducted in the hope of gaining further insight into the nature of their interrelations. Two factors emerged with eigenvalues greater than one; together, these factors accounted for 61.4% of the variance. The factor loadings, following a varimax rotation, are displayed in Table 2. The level of modern racism and the unobtrusive estimates clearly loaded on different factors, thus reiterating the independence of modern racism and automatically activated evaluations noted earlier. Consistent with our observations from the correlation matrix, the two factors clearly reflect variables that are related to level of modern racism versus those related to the unobtrusive estimates. The interaction rating loaded much more highly on the unobtrusive estimate factor, whereas the Rodney King verdict variable and the attractiveness variable loaded more highly on the modern racism factor. The responsibility variable loaded about equally on both factors. (These two factors receive further discussion later in the context of the findings of the subsequent studies.)

Discussion

The present data are certainly encouraging for what they imply about our unobtrusive measure of attitude. Two findings, in particular, suggest that the measure has some validity. First, the attitude estimates for Black and White students differed markedly. Black and White students displayed very different patterns of facilitation in response to primes that consisted of Black versus White faces. Whereas White students exhibited more negativity in response to Black faces than in response to White faces, Black students displayed the opposite pattern of relatively more negativity toward Whites.

Second, the variability that existed among White students appears meaningful in the sense that it was predictive of at least some race-related judgments and behaviors. Although the attitude estimates for White students were, on average, indicative of negativity toward Blacks, some students displayed more negativity than others. These attitude estimates related significantly to the Black experimenter's ratings of the quality of the interaction that she had with the student. Those students with more negative attitudes toward Blacks, according to the unobtrusive measure, behaved in a less friendly and less interested manner. The measure also proved predictive of the extent to which students assigned responsibility to Blacks versus Whites for the riots in Los Angeles after the Rodney King verdict. In summary, there are a number of indications that the indirect estimates obtained from the pattern of facilitation evident on trials involving positive versus negative adjectives form a valid, unobtrusive measure of attitudes toward Blacks.

These findings are of theoretical, as well as methodological, import. The data illustrate the importance of evaluations that are activated automatically from memory. The meaningful variability that we observed among the White students suggests that a personal evaluation, rather than a socially shared cultural stereotype, is automatically activated on encountering a Black person. Thus, at a theoretical level, the data point to the importance of considering the variability that exists in the automatic component of prejudice.

In addition, the research provides corroboration for the MODE model's (Fazio, 1990) depiction of a spontaneous attitude-to-behavior process and, in particular, for the emphasis that it places on evaluations that are activated from memory

Table 2
Factor Loadings

Variable	Factor 1	Factor 2
Unobtrusive estimate	−.10	.84
Modern racism level	−.83	.08
Interaction rating	.10	.68
Rodney King verdict	.84	.10
Responsibility	.59	.55
Attractiveness	.66	.05

automatically. To our knowledge, our consideration of the relation between our unobtrusive measure and behavior toward the Black experimenter represents the very first effort to directly examine the correspondence between automatically activated evaluations and behavior. Although numerous investigations have examined how attitude–behavior consistency varies as a function of attitude accessibility (i.e., the likelihood that the attitude will be activated from memory automatically on encountering the attitude object; see Fazio, 1995, for a review), the present investigation is the first to directly assess the nature of the evaluations that are automatically activated from memory and to use the resulting individual-differences measure as a predictor of behavior.

This is not to say that motivations to control the influences of any negativity that is automatically activated, in the manner that Devine (1989b) has proposed, are irrelevant. It is certainly reasonable to believe that some individuals may experience such automatically activated negativity toward Blacks but work to counter the influence of such negativity so as not to appear prejudiced to themselves or others, or both. As Devine has noted, change from the status of racially prejudiced to unprejudiced may involve precisely such a stage of attempting to consciously monitor and control one's judgments and behaviors (Devine, 1989a; Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, Devine, & Zuwerink, 1993). Some such judgments and behavior may be more difficult to control fully than others. It is for such relatively uncontrollable classes of behavior that the effects of any automatically activated personal evaluations are likely to be most apparent.

The single perplexing aspect of the findings from the present investigation was the lack of any apparent relation between our unobtrusive measure and what is probably the most commonly used self-report measure of racial prejudice, namely, the Modern Racism Scale. At the very outset of the investigation, we had been concerned about the relatively unprejudiced nature of a college student population. Indeed, the distribution of scores on the Modern Racism Scale was heavily skewed in our mass survey of nearly 500 students. Relatively few scores fell at the prejudiced end of the scale. It was for this reason that the students who participated in the study were selected from the top and bottom 10% of the distribution. The low-prejudiced students all scored at the endpoint of the scale (−14). The high-prejudiced students scored at 2 or higher.[4] Despite having sampled two groups whose scores differed so markedly, we found no signs of an effect of modern racism level on our facilitation data. The basic Race of Photo × Valence of Adjective interaction was not moderated by modern racism level. Moreover, our unobtrusive measure and modern racism loaded on two distinct factors.

In summary, the present findings suggest that the priming methodology may yield a valid, unobtrusive measure of racial attitudes. However, the discordance between our unobtrusive estimates and scores on the Modern Racism Scale is puzzling. If the unobtrusive measure is as valid as the preliminary findings lead us to believe and if the Modern Racism Scale is characterized by the validity with which it has been portrayed, then it is reasonable to expect the two measures to correlate with one another. This is the puzzle to which we turned our attention in the next three studies.

The first issue that required attention was whether the null relation observed in Study 1 would replicate. Study 2 attempted such a replication. Moreover, the participants in Study 2 were recruited with different criteria in mind. Study 1 had involved students from the top or bottom 10% of the distribution of Modern Racism Scale scores. In the case of the nonprejudiced group, the students all had the most extreme score possible on the scale (i.e., −14). Thus, they had a response of −2 to each of the seven items on the scale. Although it does not seem very likely, one might argue that there is something unusual about people who consistently use the endpoint of the scale. In addition, the lack of variability within the nonprejudiced group made it impossible to meaningfully examine the relation between our unobtrusive estimates and each of the items of the Modern Racism Scale individually. For these reasons, Study 2 involved a sample of participants recruited so as to form roughly a normal distribution of scores on the Modern Racism Scale.

Study 2

Method

Participants. Forty-nine students, all of whom were White, participated in the study. They were selected from a large sample of students who had completed the Modern Racism Scale as part of a mass survey early in the semester. The students were recruited with the goal of obtaining a roughly normal distribution of scores on the Modern Racism Scale centered around the scale value of 0. This effort was successful and resulted in a sample with a mean score of 0.33 and a standard deviation of 5.11.

Procedure. The stimulus materials and procedure for Study 2 were identical to those used in Study 1.

Results

Facilitation scores, as well as attitude estimates based on the facilitation data, were calculated in the same manner as for Study 1. These attitude estimates, for which lower scores reflect greater negativity toward Blacks, averaged −.14, which was significantly different from zero, $t(48) = 3.27$, $p < .002$. Just as in Study 1, then, students displayed relatively greater facilitation when positive adjectives had been preceded by White than by Black primes and when negative adjectives had been preceded by Black than by White primes.

The attitude estimates were correlated with scores on the Modern Racism Scale, for which higher scores are presumed to reflect greater prejudice. The correlation was .28 ($p < .06$). Although the coefficient did not reach a conventional level of statistical significance, this correlation was in the direction opposite to what was expected. People with less prejudiced scores on the Modern Racism Scale exhibited more negativity toward Blacks on our measure of automatically activated evaluations. That is, more negative scores on the unobtrusive measure tended to be associated with lower, less prejudiced scores on the

[4] It was this selection criterion that led to the serendipitous sampling of so many Blacks in our sample. Eight of our 53 students were Blacks, which is much higher than the percentage of Black students on the Bloomington campus. Seven of these 8 Blacks had scores of −14.

Modern Racism Scale. When the correlations between the unobtrusive measure and each of the seven scale items were examined individually, none were found to yield a significant correlation in the expected direction; six, three of which were statistically reliable, produced a correlation in the unexpected direction. Thus, just like in Study 1, the data from Study 2 failed to reveal the expected relation between attitude estimates based on the priming technique and Modern Racism Scale scores.

Discussion

Essentially, then, the second study substantiates what was observed in the first. Even when a normally distributed sample was selected, the expected relation between Modern Racism Scale scores and our unobtrusive estimates based on automatically activated evaluations was not observed. What, then, is the Modern Racism Scale measuring?

We have focused our attention on two potential difficulties with the Modern Racism Scale as the basis for the discordance between the scale and our unobtrusive measure. One issue concerns a confounding variable, and the other concerns the potential reactivity of the scale. First, the scale appears to confound prejudice and political conservatism. Recall that the wording of the items was intended to permit an individual's giving the prejudiced response to be "explained by racially neutral ideology" (McConahay, 1986, p. 100). The racially neutral ideology here is conservatism. Given the nature of the items (e.g., "Over the past few years, the government and news media have shown more respect to Blacks than they deserve" and "Over the past few years, Blacks have gotten more economically than they deserve"), it would appear quite difficult for a political conservative—one who does not value government intervention—to score at the low-prejudiced end of the scale. This confounding has been emphasized by critics of the modern racism perspective, especially Sniderman and Tetlock (1986a, 1986b). In fact, empirical work has documented the existence of a correlation between individuals' self-identification as politically conservative and scores on the Modern Racism Scale (Weigel & Howes, 1985). Yet, experimental results have demonstrated that political conservatives, although less supportive than liberals of governmental assistance of any form, are, if anything, more supportive of such assistance for Black claimants than for White claimants (Sniderman & Piazza, 1993; Sniderman, Piazza, Tetlock, & Kendrick, 1991).

Thus, there appears to be no reason to believe that politically conservative individuals are necessarily racially prejudiced. Nevertheless, such a confounding is implied by Weigel and Howes's (1985) finding and also is readily apparent within our own studies. In our work, the Modern Racism Scale items had been embedded in an opinion survey. Many of these filler items concerned political issues and were such that a liberal versus conservative side to the issue could be identified. Thirteen of these filler items served as our measure of conservatism. They covered a wide range of topics (e.g., abortion rights, capital punishment, gay rights, censorship, tax increases, defense spending, and sex education) but yielded a scale with quite satisfactory internal consistency (Cronbach's $\alpha = .72$). This measure of political conservatism was correlated with Modern Racism Scale scores. When our two samples of White students were combined, the correlation was highly significant ($r = .47, p < .001$), with more conservative individuals having Modern Racism Scale scores presumably indicative of greater prejudice. Thus, it seems very possible for someone to obtain a high score on the Modern Racism Scale not because he or she is prejudiced but simply because he or she is conservative.[5]

The second difficulty with the Modern Racism Scale concerns its potential reactivity. Despite the claims to the contrary, the scale items appear very obvious and even blatant. The evidence that has been marshalled in support of the scale's nonreactivity comes from research by McConahay, Hardee, and Batts (1981). These researchers had participants complete the modern racism items, as well as the so-called old-fashioned racism items. The latter included such even more blatant and inflammatory statements as "Black people are generally not as smart as Whites" and "It was wrong for the U.S. Supreme Court to outlaw segregation in its 1954 decision." The questionnaire was administered by either a Black or a White experimenter. In each of two experiments, scores on the old-fashioned items, but not the modern items, were influenced by the race of the experimenter. In a third experiment, participants rated the extent to which agreement (or disagreement) with a given item was indicative of a negative attitude toward Blacks. The old-fashioned items were perceived as more indicative of negativity than were the modern racism items, although even the latter were viewed as having significant racial implications relative to filler items that bore no relation to issues of race. On the basis of these findings, McConahay et al. (1981) concluded that

> these experiments demonstrate that the modern racism scale is a nonreactive measure in this period of history. When there were incentives to fake being less prejudiced, items measuring old-fashioned racism were the ones that subjects could spot and figure out how to fake in a consistently less prejudiced direction. Though the modern racism items might have been recognized as having racial implications, they were not altered very much or very consistently. (p. 577)

It is our belief that the participants' recognition of the racial implications of the modern racism items points to a serious problem in and of itself. The participants in McConahay et al.'s (1981) Experiment 3 perceived agreement with the modern racism items to be indicative of a negative attitude. The failure of participants in the other two experiments to shift toward a less prejudiced direction when the questionnaire was administered by a Black experimenter may bear little relation to the potential reactivity of the scale. They may not have shifted because they may have believed that the attitudes they were expressing were such that they would not offend the Black experimenter. Indeed, the participants were not very prejudiced, according to their Modern Racism Scale scores. At the time, the scale included only six items, ranging from *strongly disagree* (−2) to *strongly agree* (2); possible total scores ranged from −12 (least prejudiced) to 12 (most prejudiced). The mean in

[5] In this vein, it is interesting to note that the single Black student from Experiment 1 whose Modern Racism Scale score led to inclusion in the high-prejudice group had the second highest conservatism score in the entire sample.

the White experimenter condition was −4.2, on the nonprejudiced side of the neutral point. Thus, these participants may have believed that they held the "correct" attitudes on the modern racism items and, hence, felt little pressure to alter their responses in the presence of a Black experimenter. In contrast, they may not have perceived their slight disagreement with the even more blatant, old-fashioned items as sufficiently extreme to satisfy the Black experimenter.

Whether it be for this reason or simply because times have changed since the McConahay et al. (1981) study was conducted, we thought that it would be useful to conduct an experimental test of the reactivity of the Modern Racism Scale. All of our students had participated in a mass survey that included the Modern Racism Scale early in the semester. As a means of ensuring our sample might have reason to modify their responses, all of the students who participated in the study had scores of at least −1 on the current (−14 to 14) version of the Modern Racism Scale. Thus, our least prejudiced score was considerably more prejudiced than the average score in the McConahay et al. study. The students were administered an opinion survey that included the modern racism items by either a Black or a White experimenter.

Study 3

Method

Participants. Fifty-eight students participated for payment or partial fulfillment of an introductory psychology course requirement. These individuals were among 552 students who had participated in an earlier mass survey that included the Modern Racism Scale. Students were selected on the basis of their scores on the Modern Racism Scale; those with the highest scores (i.e., high prejudiced) were selected (scores ranged from −1 to 7). Pairs of students were created by matching students with identical scores on the Modern Racism Scale. One member of each pair was then randomly assigned to the Black or the White experimenter condition.

Procedure. Students were recruited for participation in the experiment 2 to 3 months after the mass survey. On their arrival, students were greeted by a Black or a White female experimenter unaware of their scores on the Modern Racism Scale.

Students were told that they would be completing three short questionnaires. The first questionnaire included the Modern Racism Scale; the other two measures were not related to the present concerns. Each experimenter stressed that because the questionnaires were not computer administered, she would personally input the student's responses into the computer at a later time. This statement was intended to underscore that the experimenter would be aware of the student's responses to questionnaire items.

The format of the Modern Racism Scale was modified from the format used during the earlier mass survey to minimize the possibility that students would recognize the scale, but the wording and order of the modern racism items were not changed. As before, the items were embedded in a larger opinion survey. However, new filler items were used, the questionnaire was printed on paper of a different color, and the format of the 5-point scale was altered.

Results

The mean Modern Racism Scale scores in the two conditions are presented in Figure 3. Naturally, given our creation of matched pairs, the scores in the two conditions were identical when considering the scale completed in the mass survey. However, at Time 2, the scores shifted toward the less prejudiced end of the scale. This was true of both the White experimenter condition, $t(28) = 2.21$, $p < .04$, and the Black experimenter condition, $t(28) = 6.34$, $p < .001$. However, as is apparent in Figure 3, the movement toward the less prejudiced end of the scale was especially pronounced when the scale was administered by the Black experimenter. An ANOVA revealed the Time × Race of Experimenter interaction to be highly significant, $F(1, 56) = 18.19$, $p < .001$.

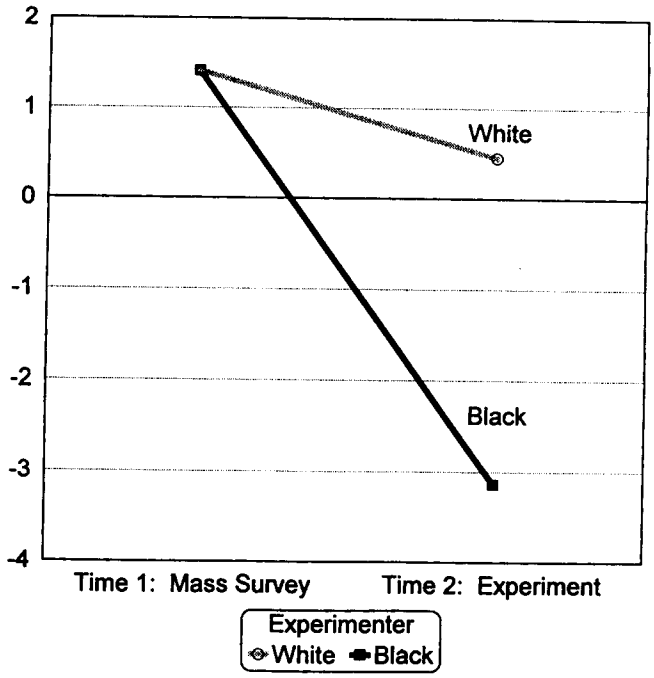


Figure 3. Mean scores on the Modern Racism Scale at Time 1 and Time 2 in the White and in the Black experimenter conditions. Higher numbers indicate more prejudice.

Additional analyses revealed that the movement induced by the Black experimenter did not simply shift the entire distribution in a less prejudiced direction. Instead, the relative rank ordering of the students was affected. The correlation between Modern Racism Scale scores obtained in the initial mass survey and the scores obtained at Time 2 was .68 ($p < .001$) when the Time 2 experimenter was White and .29 ($p = .12$) when she was Black. A test of the difference between the two correlations revealed a marginally significant difference ($z = 1.93$, $p < .054$).

Discussion

On the basis of the present findings, it certainly appears that the Modern Racism Scale is a reactive measure. Students did indeed respond in a less prejudiced manner when interacting with the Black experimenter. Apparently, students were not only aware of the racial implications of the scale items but also motivated to respond in a less prejudiced manner as a result. Thus, the experiment provides evidence that the Modern Racism Scale, like traditional self-report measures, is subject to social desirability concerns.

How might one account for the difference between our results and those of McConahay et al. (1981)? Students in the present sample had average Modern Racism Scale scores that placed them on the prejudiced side of the scale. In contrast, McConahay et al.'s participants had scores that placed them on the nonprejudiced side of the scale. Thus, participants in the present study may have been more motivated to shift toward the nonprejudiced end of the scale than those in McConahay et al.'s study. Participants in McConahay et al.'s (1981) study may not have felt this same pressure because their responses were sufficiently nonprejudiced to raise little concern about social desirability.

Alternatively, or in addition, the difference between the outcomes of the two studies may simply reflect the possibility that the scale items have become outdated, a possibility that McConahay et al. (1981) seem to have forecast when they concluded that their findings indicated the Modern Racism Scale to be a "nonreactive measure in this period of history" (p. 577). Although McConahay et al.'s (1981) participants were aware of the racial implications of the scale items, they may not have labeled these items as blatantly racist. However, statements that were perceived as subtle and nonracist more than a decade ago may now be perceived as blatantly racist. Indeed, recent research by Swim, Aikin, Hall, and Hunter (1995) suggests that old-fashioned and modern racism scale items are no longer as empirically distinct as they once were. Whether it was because we selected a sample with relatively prejudiced Modern Racism Scale scores or simply because times have changed, the present data clearly contradict the conclusion that the Modern Racism Scale is nonreactive.

It should be emphasized that the Black–White experimenter manipulation merely provides a convenient and effective means of examining the scale's inherent reactivity. It by no means suggests that scores on the Modern Racism Scale are valid with a White experimenter. In our view, social desirability concerns are accentuated when it is evident that a Black experimenter will see the participant's questionnaire. However, the extent to which such factors concern a given individual and the extent to which the concerned individual adjusts his or her responses varies considerably, as is suggested by the differential correlations of Modern Racism Scores across time in the two conditions. Moreover, these concerns are likely to arise for some individuals even when the scale is distributed by a White experimenter (as it was during the initial mass survey), simply as a consequence of the blatantly obvious wording of the scale items. Were it not for the language of the Modern Racism Scale items, the race of the experimenter would matter little. As a result, some truly prejudiced individuals obtain scores presumably indicative of low prejudice, no matter how the scale is administered.

In this context, it is interesting to speculate about the relation that was observed in Study 2 between Modern Racism Scale scores and our unobtrusive estimates of attitude. Recall that the correlation was in the unexpected direction; students for whom more negativity was automatically activated in response to Black faces had lower, less prejudiced scores on the scale. The inherent reactivity of the Modern Racism Scale may have prompted students for whom negative attitudes were automatically activated to present themselves as relatively unprejudiced on the scale items. When faced with such social desirability concerns, these students may have, in effect, overcompensated (i.e., presented themselves as even less prejudiced than individuals for whom little or no negativity toward Blacks is automatically activated).

As noted earlier, another force that contributes to the manner in which respondents complete the Modern Racism Scale is their political conservatism. The Modern Racism Scale appears to be confounded with conservatism. Thus, some truly nonprejudiced individuals obtain scale scores presumably indicative of prejudice simply because they are conservative. It is our contention that these two forces—the scale's reactivity and its confounding with conservatism—yield a distribution of Modern Racism Scale scores that bears little resemblance to the extent to which negative evaluations are automatically activated on encountering a Black individual. These arguments suggest that the Modern Racism Scale may amount to a measure of "willingness to express" negativity toward Blacks that is confounded with political conservatism. This reasoning was tested in Study 4.

Study 4

According to the MODE model (Fazio, 1990), attitudes can influence judgments and behavior via either a spontaneous or a deliberative process. However, the model explicitly postulates the possibility of attitude–behavior processes that are neither purely spontaneous nor purely deliberative but instead are "mixed" processes involving a combination of automatic and controlled components. The MODE model also proposes that a deliberative process, or any controlled component within a mixed sequence, requires that the individual both be motivated to engage in the necessary cognitive effort and have the opportunity to do so.

It is this theoretical perspective that underlies our reasoning regarding individuals' responses to the Modern Racism Scale. Any automatically activated negativity toward Blacks is potentially tempered by a controlled process among individuals who are so motivated. The reactivity that was observed in Study 3 suggests that at least some individuals experience such motivation when responding to the items of the Modern Racism Scale. In Study 4, we directly assessed this motivation to control seemingly prejudiced reactions. Scores on the Modern Racism Scale were predicted to be a function of political conservatism and the joint influence of automatically activated evaluations and the motivation to control prejudice. Thus, when conservatism is controlled, Modern Racism Scale scores should be more predictable from our unobtrusive estimates of automatically activated attitudes as the motivation to control seemingly prejudiced reactions decreases.

Method

Participants. During each of two consecutive semesters, students were selected from a large sample of those who had completed the Modern Racism Scale as part of a mass survey early in the semester. One hundred seventeen students were recruited with the goal of obtaining a wide range of scores on the Modern Racism Scale. The sample's mean score on the scale was −2.33, with a standard deviation of 7.39.

Procedure. The same experimenter conducted the priming procedure in each of the two semesters. Students participated in the same first

five procedural tasks used in Study 1, with the exception that 16 Black faces and 16 White faces were paired for the target trials. The remaining 16 photos (Whites, Hispanics, and Asians) were used on the filler trials. Given time constraints and the lack of any need to set the stage for a debriefing by a Black experimenter, attractiveness ratings were not collected in Study 4.

Motivation to control racial prejudice. Students completed a 17-item scale assessing their motivation to control seemingly prejudiced reactions. The survey included such statements as "In today's society it's important that one not be perceived as prejudiced in any manner"; "If I have a prejudiced thought or feeling, I keep it to myself"; and "It's never acceptable to express one's prejudices." Students responded to each item on a scale ranging from *strongly disagree* (−3) to *strongly agree* (3). The scale's internal consistency within the present sample was more than adequate (Cronbach's $\alpha = .81$). All of the scale items, as well as additional psychometric details, have been reported by Dunton and Fazio (1995).

Students in one semester completed the scale as their final task in an ostensibly unrelated second session conducted by a different experimenter 1 week after the priming procedure. In the other semester, the scale was included as part of the mass survey.

Political conservatism. As in Studies 1 and 2, political conservatism was estimated by students' responses to relevant filler items in the opinion survey in which the Modern Racism Scale items had been embedded. As before, the internal consistency of this scale was satisfactory (Cronbach's $\alpha = .68$).

Results

Estimates of automatically activated attitudes were calculated from facilitation scores in the same manner as in Studies 1 and 2. Although not as extremely as in the earlier studies, the average attitude estimate tended to be negative ($M = -.046$), $t(116) = 1.74$, $p = .084$. Moreover, there was considerable variability around this mean ($SD = .287$).

Complete data (i.e., unobtrusive estimates, motivation to control prejudice scores, and conservatism scores) were available for 111 of the 117 students. The major hypothesis was tested by a multiple regression equation predicting scores on the Modern Racism Scale from conservatism, attitude estimates, motivation scores, and the interaction between the last two variables. When entered simultaneously, the three main effects accounted for a significant proportion of the variance, $R = .57$, $F(3, 107) = 17.57$, $p < .001$. Conservatism was a highly significant predictor, $t(107) = 6.90$, $p < .001$, such that more politically conservative individuals had higher, more prejudiced Modern Racism Scale scores. This finding replicates what was observed earlier in Studies 1 and 2. Higher motivation to control prejudice was significantly associated with lower, less prejudiced scores on the Modern Racism Scale, $t(107) = 2.68$, $p < .01$. No main effect of the attitude estimate was apparent, $t < 1$.

The interaction between attitude estimates and motivation to control prejudice was entered as the next step in the equation and yielded a significant increase of .03 in the squared multiple correlation, $F(1, 106) = 5.38$, $p < .025$. Other interaction terms were tested but were not found to add to the prediction of Modern Racism Scale scores. Nor did the semester in which the students participated yield a main effect or any interactions.

The nature of the Attitude × Motivation interaction can be readily discerned from Figure 4, which displays the regression lines predicting Modern Racism Scale scores from attitude estimates for motivation to control prejudice scores of 1.07 and −0.59 (one standard deviation above and below the mean of 0.24). As motivation to control prejudice decreases, the relation between the unobtrusive attitude estimates and Modern Racism Scale scores grows stronger, such that more negative automatically activated attitudes are associated with more prejudiced Modern Racism Scale scores. In other words, motivation to control prejudice matters little among those for whom little or no negativity is automatically activated in response to Black faces. However, motivation to control seemingly prejudiced reactions exerts a strong influence among those individuals for whom negativity is automatically activated. Those with little such motivation feel free to respond to the Modern Racism Scale items in a manner that is indicative of prejudice, whereas more motivated individuals present themselves as far less prejudiced.

Discussion

The results of Study 4 confirm the predictions advanced by the MODE model and our reasoning regarding the Modern Racism Scale, as well as attesting further to the predictive validity of our unobtrusive estimates of racial attitude. As suggested earlier, the Modern Racism Scale assesses willingness to express prejudice and conservatism. Unprejudiced individuals can score high on the scale if they endorse a conservative political philosophy. Truly prejudiced individuals will score high on the scale, provided that they are not motivated to control prejudiced reactions. Relatively low scores on the Modern Racism Scale, on the other hand, can emanate either from individuals' being truly unprejudiced or from their being motivated to control their prejudiced reactions.

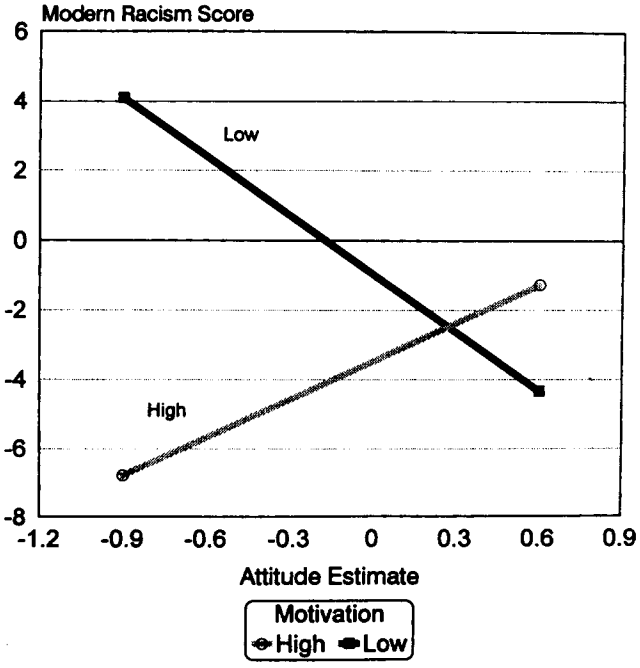


Figure 4. Regression lines predicting scores on the Modern Racism Scale as a function of the unobtrusive attitude estimates and motivation to control prejudice. Higher scores indicate more prejudice.

Obviously, individuals can control their responses to the Modern Racism Scale items if they are motivated to do so. The interaction that was observed between automatically activated attitudes and motivation indicates that individuals who lack such motivation respond in a manner that is consistent with their automatically activated attitudes. In contrast, individuals who are motivated to control prejudiced reactions do not respond in accord with their automatically activated attitudes. Indeed, the positive slope that was observed among those with high motivation suggests that individuals who are highly motivated to control their automatically activated negativity may actually overcompensate. They presented themselves as even less prejudiced than similarly motivated individuals for whom negativity is not automatically activated. This observation is consistent with our earlier conjecture regarding the positive correlation between Modern Racism Scale scores and our unobtrusive estimates in Study 2, a correlation that was in the direction opposite to what was expected. Apparently, a preponderance of the students in Study 2 were relatively motivated to control prejudiced reactions.

The findings certainly suggest that researchers need to be extremely cautious in using the Modern Racism Scale. Given the myriad of factors that influence individuals' responses to the scale items, it will be difficult to draw any clear inferences from findings based on the scale scores. This is not to say that the Modern Racism Scale is not without predictive validity. Indeed, it was found to be predictive of some race-related judgments in Study 1. Given that the Modern Racism Scale measures willingness to express negativity toward Blacks, along with conservatism, it should be predictive of any judgments that themselves involve either or both of these dimensions. Thus, for example, it should correlate reasonably well with any judgment that is itself influenced by conservatism.[6] Likewise, it should correlate reasonably well with any judgment that provokes the same concerns with social sensitivity. In this case, individuals who are willing to express negativity on the Modern Racism Scale are equally likely to be willing to express negativity on some other clearly race-related judgment.

General Discussion

The major aim of the present research concerned the value of using the priming paradigm that has been used to study automatic attitude activation as an unobtrusive measure of racial attitudes. The priming procedure appears to provide a bona fide pipeline for attitude measurement. It permits assessment of the extent to which judgments of positive versus negative adjectives are facilitated by primes consisting of Black faces relative to primes consisting of White faces. As noted earlier, the findings of Study 1 suggest that attitude estimates derived from the resulting patterns of facilitation have some validity. Similarly, the results of Study 4 indicate that the unobtrusive attitude estimates are predictive of responses to the Modern Racism Scale for individuals who are relatively unmotivated to control any prejudiced reactions that they may experience.

The results of the research clearly attest to the importance of the variability that exists among people with respect to the evaluations that are likely to be activated from memory automatically on encountering a minority individual. The theoretical assumption offered by Devine (1989a, 1989b) concerning the automatic activation of a socially shared, cultural stereotype involving negativity toward Blacks does not appear tenable in light of the present results. Nevertheless, the findings, especially those from Study 4, imply that it may be useful to identify, broadly speaking, three types of individuals, one of whom is very appropriately characterized by Devine's model. The three types vary with respect to the evaluation that is automatically activated and with respect to the extent to which the evaluation is countered by a subsequent controlled process. First, some individuals do not experience the automatic activation of any negative evaluation from memory on encountering a Black person (or may experience activation of a positive evaluation relative to what occurs when they encounter a White target). We would label these individuals as truly nonprejudiced. Second, negativity is automatically activated for other individuals. Some such people may have no qualms about their experiencing such negativity or about expressing it. It seems appropriate to label these individuals as truly prejudiced. Third, as Devine (1989a, 1989b) has postulated, some individuals for whom negativity is automatically activated may be motivated to counter the effects of that negativity.[7] This motivation may vary from a sincere distaste for the negative reaction that was automatically evoked on encountering a Black individual to a more strategic self-presentation dictated by perceptions of the social norms for the particular situation in which the individual was encountered. In either case, the expression of judgments and the performance of overt behavior may be carefully and deliberately monitored so as to avoid the appearance of a prejudiced response. The more the efforts to inhibit and control one's automatically activated negativity stem from a sincere displeasure with one's having experienced such negativity, the more one would appear to be moving toward a truly nonprejudiced stand. The importance of such self-dissatisfaction has been documented by recent research concerned with the process of prejudice reduction (Devine et al., 1991; Monteith, 1993; Monteith et al., 1993).

We believe that the MODE model (Fazio, 1990) provides a useful theoretical framework for considering the influence of automatic and controlled processes in race-related judgments and behavior (see Dovidio & Fazio, 1992). As noted earlier, the model asserts that motivation is necessary to engage in the cognitive effort required to counter any automatically activated

[6] In fact, conservatism scores did correlate significantly with judgments of the Rodney King verdict ($r = .35$, $p < .025$). More conservative individuals viewed the verdict as more appropriate, which may have contributed to the relation observed between modern racism level and verdict judgments.

[7] A fourth class of individuals, one that can be viewed as a subcategory of the second class, may be identifiable. As noted earlier, some individuals may have negative evaluations of Blacks but not be aware of their sentiments or experience any automatically activated negativity phenomenologically (see Greenwald & Banaji, 1995). Nevertheless, such individuals' behavior would presumably be influenced by their automatically activated attitudes. Yet, they may not realize that there is any need for them to monitor and control their behavior. In this sense, such individuals are similar to individuals who are aware of their negativity but have no reservations about expressing it. For both such types, the automatically activated negativity is not tempered by a controlled process.

negativity. In this particular domain, the relevant motivation concerns a desire to control and avoid seemingly prejudiced responses. Whether this motivation is engaged in any given situation will depend on individuals' construal of the situation and the judgmental or behavioral response that is being requested of them. Construing the situation and requested response as related to their racial attitudes will evoke such motivation among any individuals who share these concerns.

The interactive influence of automatically activated attitudes and motivation to control prejudice that characterized responses to the Modern Racism Scale in Study 4 presumably stemmed from such construals. In fact, the reactivity that was demonstrated in Study 3 illustrates that the scale items are perceived as related to racial attitudes. We would suggest that any requested judgment or behavior that is construed as related to racial attitudes will evoke motivational concerns. Study 1's assessments of the Rodney King verdict and of the attractiveness of the Black and White faces, both of which were found to load highly on the same factor as the Modern Racism Scale, are likely to have been of this sort. To some extent, the same may have been true of Study 1's assessment of Black versus White responsibility for the riots that followed announcement of the verdict, a measure that loaded highly on both the modern racism and the unobtrusive estimate factors.

According to the MODE model, the opportunity factor will be the key determinant of whether any motivation that is activated can successfully counter the influence of automatically activated negativity. As mentioned earlier, efforts to control any automatically activated negativity may vary in their likelihood of success. Some behaviors may be more easily monitored and controlled than others. Likewise, some situations (e.g., those that do not place the individual under any time pressure to respond) may offer a greater opportunity for the motivated individual to engage in the desired controlling effort than other situations (Jamieson & Zanna, 1989; Kruglanski & Freund, 1983; Sanbonmatsu & Fazio, 1990). Judgments and behaviors that provide the opportunity for controlling one's response should be predictable from the joint, interactive influence of automatically activated attitudes and motivation to control prejudice, just as we observed for the Modern Racism Scale. Such controllable responses also are likely to be well predicted from other measures, like the Modern Racism Scale, that are similarly influenced by the extent to which the individual is willing to express a response that is identifiable as prejudiced. However, as opportunity decreases, either because the behavior is not easily controllable or because the situation itself limits the opportunity, the behavior should be less influenced by motivational concerns and more directly influenced by any automatically activated evaluations. Thus, as the opportunity to satisfy one's motivation to control prejudice decreases, the behavior should be more and more predictable from attitude estimates based on automatically activated evaluations.

Nonverbal behavior, in particular, may be subject to "leakage" of the negativity that an individual is experiencing, despite the individual's effort to behave in a nonprejudiced manner (see Ambady & Rosenthal, 1992; Rosenthal & DePaulo, 1979). Indeed, one of the most interesting of our findings may be the ability of our unobtrusive measure to predict the Black experimenter's judgment of the quality of the interaction she had with the student. Apparently, those students for whom negative evaluations were automatically activated on the presentation of a Black face behaved more distantly during their interaction with her. Given what is known about nonverbal behavior and leakage, this automatically activated negativity would be difficult for such an individual to conceal during an interaction with a Black. Such leakage may have played a role in the correspondence that was observed between the unobtrusive attitude estimates and the experimenter's ratings of the quality of the interaction. Obviously, this finding requires replication before any conclusions can be reached with confidence. Yet, the finding does suggest that our unobtrusive measure can directly predict behavior, especially when the behavior measure is one that is likely to be difficult to control strategically and, hence, is influenced largely by automatically activated attitudes unmoderated by any motivation to control prejudice. We look forward to future research pursuing these possibilities in a systematic and detailed fashion.

References

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behaviors as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 256–274.

Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. Zanna & J. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7, pp. 55–76). Hillsdale, NJ: Erlbaum.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68,* 181–198.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic activation effect. *Journal of Personality and Social Psychology, 62,* 893–912.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin, 87,* 546–563.

Devine, P. G. (1989a). Automatic and controlled processes in prejudice: The role of stereotypes and personal beliefs. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 181–212). Hillsdale, NJ: Erlbaum.

Devine, P. G. (1989b). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60,* 817–830.

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.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.

Dovidio, J. F., & Gaertner, S. L. (1991). Changes in the nature and expression of racial prejudice. In H. Knopke, J. Norrell, & R. Rogers (Eds.), *Opening doors: An appraisal of race relations in contemporary America* (pp. 201–241). Tuscaloosa: University of Alabama Press.

Dunton, B. C., & Fazio, R. H. (1995). *An individual difference measure of motivation to control prejudiced reactions.* Unpublished manuscript, Indiana University, Bloomington.

Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & E. T. Higgins (Eds.), *The handbook of motivation and cognition* (pp. 204–243). New York: Guilford Press.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology.* (Vol. 23, pp. 75–109). New York: Academic Press.

Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation: Data re-analysis and commentary on Bargh, Chaiken, Govender, and Pratto (1992). *Journal of Personality and Social Psychology, 64,* 753–758, 764–765.

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

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

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando, FL: Academic Press.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23–30.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–517.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion, 8,* 515–533.

Jackson, J. R., & Fazio, R. H. (1994, May). *Assessment of racial attitudes.* Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.

Jackson, J. R., & Fazio, R. H. (1995, May). *The effect of race of experimenter on responses to the Modern Racism Scale: A test of reactivity.* Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.

Jamieson, D. W., & Zanna, M. P. (1989). Need for structure in attitude formation and expression. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 383–406). Hillsdale, NJ: Erlbaum.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76,* 349–364.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19,* 448–468.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution, 25,* 563–579.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65,* 469–485.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology, 64,* 198–210.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). "Us" and "them": Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology, 59,* 475–486.

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26,* 199–216.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Rosenthal, R., & DePaulo, B. M. (1979). Sex differences in accommodation in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences* (pp. 68–103). Cambridge, MA: Oelgeschlager, Gunn, & Hain.

Sanbonmatsu, D. M., & Fazio, R. H. (1986, October). *The automatic activation of attitudes toward products.* Paper presented at the meeting of the Association for Consumer Research, Toronto, Ontario, Canada.

Sanbonmatsu, D. M., & Fazio, R. H. (1990). The role of attitudes in memory-based decision making. *Journal of Personality and Social Psychology, 59,* 614–622.

Sanbonmatsu, D. M., Osborne, R. E., & Fazio, R. H. (1986, May). *The measurement of automatic attitude activation.* Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, IL.

Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology, 19,* 247–255.

Sniderman, P. M., & Piazza, T. (1993). *The scar of race.* Cambridge, MA: Harvard University Press.

Sniderman, P. M., Piazza, T., Tetlock, P. E., & Kendrick, A. (1991). The new racism. *American Journal of Political Science, 35,* 423–447.

Sniderman, P. M., & Tetlock, P. E. (1986a). Reflections on American racism. *Journal of Social Issues, 42,* 173–187.

Sniderman, P. M., & Tetlock, P. E. (1986b). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues, 42,* 129–150.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68,* 199–214.

Weigel, R. H., & Howes, P. W. (1985). Conceptions of racial prejudice: Symbolic racism reconsidered. *Journal of Social Issues, 41,* 117–138.

Received January 19, 1995
Revision received July 5, 1995
Accepted July 10, 1995 ■