Reducing Automatically Activated Racial Prejudice Through Implicit Evaluative Conditioning

Michael A. Olson
University of Tennessee

Russell H. Fazio
The Ohio State University

The authors report a set of experiments that use an implicit evaluative conditioning procedure to reduce automatically activated racial prejudice in White participants in a short period and with relatively few trials. Experiment 1 demonstrated that participants were unaware of the repeated conditioned stimulus–unconditioned stimulus (CS-US) pairings of Black-good and White-bad. In Experiment 2, the procedure was found to be effective in reducing prejudice as indicated by an evaluative priming measure of automatically activated racial attitudes. In Experiment 3, this reduction in prejudice was found to persist throughout a 2-day separation between the conditioning procedure and the administration of the dependent measure. The implications of the present findings for the persistence of automatically activated racial prejudice are discussed.

Keywords: *implicit social cognition; prejudice; attitude change; conditioning*

Attitudes toward racial groups develop through socialization processes and personal experience early in life, and by adulthood, these attitudes are thought to be deeply entrenched in the cognitive repertoire (Devine, 1989). In fact, racial prejudices are often so well learned that they are activated automatically upon encountering a member of relevant groups (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Olson & Fazio, 2003) and become the first piece of input on the path toward discriminatory behavior (e.g., Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; McConnell & Liebold, 2001). Such automatically activated attitudes guide attention toward attitude-consistent information, provide a template with which to interpret ambiguous information, and can guide behavior in a relatively spontaneous fashion (for a review, see Fazio & Olson, 2003a). Although additional information is often integrated into judgments and behavior, immediate evaluative responses to objects in one's environment often dictate one's behavior (Fazio, 1990). In the present article, we will consider the possibility of changing these responses.

In the race domain, priming measures often have been employed to assess automatic attitudes (Fazio & Olson, 2003b). These measures provide estimates of the evaluative responses that race-related primes (e.g., a Black or White face) evoke based on the ease with which they facilitate the identification of subsequently presented evaluative adjectives (e.g., *awesome*). Such estimates of automatically activated racial attitudes have been shown to relate to Whites' impressions of Black social targets (Olson & Fazio, 2004), ratings of the "typical Black male undergraduate" (Dunton & Fazio, 1997), evaluations of a Black candidate for a volunteer position (Olson & Fazio, 1999), emotional reactions to commercials aimed at increasing awareness of racism (Fazio & Hilden, 2001), and a Black experimenter's impressions of White participants (Fazio et al., 1995). These measures predict not only Whites' anticipated comfort in various interracial interaction situations (Towles-Schwen &

Authors' Note: This research was based on the first author's doctoral dissertation from Indiana University under the supervision of the second author, who provided supported through Senior Scientist Award MH01646 and Grant MH38832 from the National Institute of Mental Health. The authors thank Ed Hirt, Sarah Queller, and Jim Sherman, who served on the dissertation committee, and Julie Brown, Lindsey Cudlike, Eric Currence, Roxanne Flowers, Alexis Greenhut, Kevin Griffin, Katie Hershey, Don Ladwig, Nicole DeLand, Sarah Lifland, and Dorey Meslemani for help in data collection. Portions of this work were presented at the 2002 meeting of the Midwestern Psychological Association in Chicago, Illinois. Please address correspondence to Michael Olson, Department of Psychology, Austin Peay Building, University of Tennessee, Knoxville, TN 37996; e-mail: olson@utk.edu.

PSPB, Vol. 32 No. 4, April 2006 421-433
DOI: 10.1177/0146167205284004

Fazio, 2003) but also the success of actual Black-White roommate relationships (Towles-Schwen & Fazio, 2002). Similar measures have been used to demonstrate that negative, automatically activated attitudes toward Blacks "leak" through a number of nonverbal channels (Dovidio et al., 1997; Dovidio, Kawakami, & Gaertner, 2002).

It is only when individuals are motivated to avoid the influence of their automatically activated prejudices and have ample opportunity to make the appropriate correction that these prejudices do not guide behavior in a relatively spontaneous fashion (Fazio, 1990; Fazio & Towles-Schwen, 1999). Given people's inability to adequately detect and correct for bias (Wilson & Brekke, 1994) or monitor their nonverbal behavior (e.g., Dovidio et al., 1997), automatic prejudices are pervasive influences in everyday life.

The evidence that racial prejudices develop early and eventually become automatic has led most researchers to conclude that they are relatively difficult to change. It has been argued that a more fruitful opportunity for reducing racial discrimination occurs further "downstream" in the form of conscious control over the application of racial prejudice, and much energy has been allocated to understanding how people can prevent the manifestation of prejudice in actual behavior (Devine & Monteith, 1999; Fiske, Lin, & Neuberg, 1999; Olson & Fazio, 2004). Essentially, the aim is to achieve control over the potential consequences of what is automatically activated. Although not denying the relevance and importance of such approaches, we adopt a more direct aim in the present research: to modify the nature of the attitude that is automatically activated. In the present article, we provide evidence that automatically activated racial attitudes can be changed more readily than commonly claimed.

Altering the Automatic

Recently, a few researchers have investigated whether racial prejudice might be reduced at the automatic level. For example, in two quasi-experiments with elaborate controls, Rudman, Ashmore, and Gary (2001) demonstrated reduced automatically activated negativity toward Blacks on several implicit measures of prejudice after a semester-long prejudice seminar. While providing evidence that those supposedly stubborn automatic prejudices are somewhat amenable to change, several weeks of reading, discussion, and other classroom activities were employed. So it is unclear from this research how foreboding a challenge is involved.

Other researchers have been successful at altering the automatic component of racial prejudice with less time and effort. For example, Kawakami, Dovidio, Moll, Hermsen, and Russin (2000) presented participants with images of skinheads that were sometimes paired with stereotypic traits (e.g., "criminal") and sometimes paired with nonstereotypic traits. Participants were asked to avoid thinking about the skinhead stereotype when they saw one paired with a stereotypic term and press a button labeled "NO" and to respond by pressing "YES" whenever they saw a skinhead paired with a nonstereotypic word. After the lengthy, 480-trial training session, a Stroop task revealed an elimination of the automatic stereotype activation effect seen in an earlier pretest. Thus, it appears that repeated practice at suppressing stereotypic associations and replacing them with alternative associations can successfully undermine their automatic activation.

Caveat: Malleability of Implicit Measures

When considering research that has been discussed as indicative of the malleability of automatically activated attitudes (for a review, see Blair, 2002), it is important to ask whether the findings reflect a change in the underlying representation of the attitude object or merely a change in how the object is construed. Several studies have shown that failure to categorize outgroup targets as such prevents negativity from being activated. (e.g., Livingston & Brewer, 2002; Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997). Members of outgroups also can be flexibly categorized into more positively evaluated groups under certain circumstances (e.g., Mitchell, Nosek, & Banaji, 2003; Wittenbrink, Judd, & Park, 2001; see also Macrae, Bodenhausen, & Milne, 1995), especially if such categorizations serve the self (Sinclair & Kunda, 1999; see also Lowery, Hardin, & Sinclair, 2001). These findings are consistent with the well-accepted notion that how a multiply-categorizable object is construed will determine the attitude that is automatically activated (Fazio, 1986; Fazio & Dunton, 1997; Smith, Fazio, & Cejka, 1996). Thus, the above findings reflect "a change in the object of judgment, rather than in the judgment of the object," to use Asch's (1940) classic description of "change of meaning" (p. 458).

Other research sometimes discussed as pointing to the malleability of automatically activated attitudes may reflect the malleability of the implicit measure itself. As research has begun to illuminate the mechanisms that underlie the Implicit Association Test (IAT) (see De Houwer, Thomas, & Baeyens, 2001; Olson & Fazio, 2003), it has become evident that construal of the category labels plays a critical role. Moreover, such construals are context dependent. For example, Dasgupta and Greenwald (2001) exposed participants to admired Blacks (e.g., Martin Luther King) and disliked Whites (e.g., Charles Manson), who then showed reduced prejudice toward Blacks on an IAT (also see Blair, Ma, & Lenton, 2001). Similarly, Govan and Williams (2004)

demonstrated that participants can appear to prefer insects to flowers on an IAT if the exemplars (likeable insects like fireflies and unlikable flowers like nettles) redefine the category labels. The insect and flower exemplars chosen by Govan and Williams (2004) prompted participants to define the categories as *liked* insects and *disliked* flowers, just as the Black and White exemplars chosen by Dasgupta and Greenwald implied that the categories were *liked* Blacks and *disliked* Whites.

Research demonstrating that temporary recategorization of outgroup targets as members of liked groups reduces or eliminates the activation of prejudice says little about the underlying prejudicial associations. In these cases, the associations in memory that lead to the activation of negativity in response to outgroup members remain intact and ready to resurface in more normal contexts. Thus, this research does not provide evidence for a change in the attitude that is normally activated in response to outgroup members. Likewise, evidence of malleability of a particular attitude measure is not equivalent to evidence of attitude change. The latter implies a change in the representation in memory of the attitude and the evaluation the object evokes when it is encountered in the environment (Fazio & Olson, 2003b). It is critical that research aimed at demonstrating a reduction in prejudicial associations to outgroups rules out the possibility that any observed attitude change is due to a temporary change in the way outgroup members are categorized within the experimental context.

Goals of the Present Research

The present research involves the intersection of prejudice reduction and a particular form of learning—evaluative conditioning. As such, it stands to inform both of these domains. We described earlier how both Rudman et al. (2001) and Kawakami et al. (2000) offer some evidence that a reduction in the automatic component of prejudice is possible. One of our aims was to extend this previous work using a different influence strategy. More novelly, however, we sought to accomplish this attitude change through a nonconscious learning process. Both Rudman et al.'s (2001) and Kawakami et al.'s (2000) participants were required to put forth a considerable amount of conscious effort for the sake of altering prejudicial associations. Demonstrating a change in these automatic associations through a relatively effortless unconscious learning mechanism would suggest that these attitudes may be more readily changed than previously thought. Moreover, it also would suggest the intriguing possibility that unconscious learning may sometimes have a greater impact on attitudes than conscious learning.

Additional goals of the present research relate to evaluative conditioning (EC). The intuitively appealing idea underlying EC is that pairings of an attitude object with other valenced objects will change the attitude in the direction of the objects with which it was paired. It is often argued that by being exposed to repeated pairings of members of stigmatized groups and negative events, individuals develop negative racial attitudes capable of automatic activation. That is, racial attitudes in particular, and attitudes in general, are thought to develop through processes akin to classical conditioning (De Houwer et al., 2001; Olson & Fazio, 2001, 2002).

However, EC is not a phenomenon that is as established as our intuitions might lead us to believe. Previous demonstrations of EC (e.g., Krosnick, Betz, Jussim, Lynn, & Stephens, 1992; Razran, 1938) have been criticized on a number of grounds (Eagly & Chaiken, 1993), and the mechanism underlying EC is still poorly understood (De Houwer et al., 2001; Field, 2000; Shanks & St. John, 1994). Moreover, until very recently, there was little solid evidence that EC could occur implicitly—without awareness of the stimulus pairings. This motivated us to provide more solid evidence that EC could occur implicitly, that is, when individuals show no explicit memory for the pairings.

Capitalizing on work on implicit covariation detection in cognitive psychology (Aslin, Saffran, & Newport, 1998; Chun & Jiang, 1999), we developed a paradigm (Olson & Fazio, 2001) by which we successfully conditioned participants' attitudes toward novel objects. Participants were told that the experiment was about "attention and surveillance" and that they would see a stream of randomly assembled images on the computer screen. Their task was to press a button whenever a prespecified target item appeared. Throughout that task, critical conditioned stimulus–unconditioned stimulus (CS-US) pairs were interspersed. On a surprise evaluation task, participants evaluated the CS paired with positive items more favorably than the CS paired with negative items on both an explicit and implicit measure. A surprise covariation estimation task revealed that they were unaware of the CS-US pairings. Moreover, participants were unaware of pairings of random objects that they had seen presented together as many as 20 times throughout the experiment. This EC effect was replicated more recently using a priming measure of attitudes (Olson & Fazio, 2002).

Here we hope to utilize this EC paradigm, not as a means of influencing the formation of attitudes toward novel objects but as a social influence technique for changing Whites' automatically activated attitudes toward Blacks. Thus, another goal of the present research involves examining whether EC is an effective attitude

change technique in the case of real-world, preexisting attitudes.

Yet another aim is to investigate whether conditioning at the level of the exemplar will affect attitudes toward the category. That is, will EC occur when the stimuli require participants to detect covariations at a higher level of abstraction than is apparent in the actual CS-US pairs they view? Previous EC research typically has followed a procedure where a given object is paired with positive or negative stimuli and then evaluations of that object are later assessed. In fact, prior work has used such impoverished stimuli that distinctions between exemplars and categories were nonexistent. The CS did not vary from one presentation to another; the very same image appeared repeatedly. We used more realistic, complex images as CSs in the present studies—photos of targets who could be categorized in multiple ways (e.g., age, gender, occupation). The procedure requires participants to detect (implicitly) covariation between valence of the US and racial category through exposure to individual exemplars. It is unclear from previous EC work that individuals are capable of doing so. Yet, the cognitive science literature on implicit learning suggests that perceivers may learn such statistical regularities (e.g., Aslin et al., 1998).

In short, this work stands to achieve several intersecting goals. We hope to demonstrate a positive change in automatically activated racial attitudes that (a) rules out recategorization of outgroup members as an alternative explanation and (b) does so through an implicit learning process. In using an EC technique, we also hope to demonstrate (c) change in preexisting attitudes toward important real-world objects through EC and (d) that covariation across exemplars can affect attitudes toward the category.

METHODOLOGICAL OVERVIEW

White participants were recruited to take part in two, ostensibly unrelated studies. The first was a variant of the conditioning studies reported in Olson and Fazio (2001). Participants were told that we were interested in attention and rapid responding and that they would be viewing a series of people in various roles along with a variety of distracter objects. Their task was to press a button whenever a prespecified target appeared. Interspersed throughout the procedure were critical White-negative and Black-positive CS-US image pairs. In the control condition, participants viewed the same stimuli, including all CSs and USs, but they remained unpaired, resulting in no relationship between race and valence.

After the conditioning phase, participants were submitted to a test of their awareness of the pairings (Experiment 1), or Fazio et al.'s (1995) priming measure of racial attitudes, in a supposedly unrelated separate study. The latter was administered either shortly after the conditioning phase (Experiment 2) or after a 2-day delay (Experiment 3). We predicted participants would remain unaware of the pairings but that the priming measure would reveal more positive attitudes toward Blacks for participants in the experimental condition compared to the control condition.

EXPERIMENT 1: AWARENESS

The issue of awareness in EC research has been a consistent stumbling block in the interpretation of research findings, and most researchers agree that early demonstrations (e.g., Razran, 1938) could not rule out demand effects (De Houwer et al., 2001). Contingency awareness is relevant to the experiments proposed here for two reasons. First, skeptics are quick to point out the possibility of awareness-mediated attitude change in EC studies (e.g., Field, 2000; Shanks & St. John, 1994). Hence, it is important to provide evidence that any attitude change is not awareness-mediated. Second, prior research has considered only conscious or controlled routes to changing automatically activated racial attitudes. Unique to the experiments proposed here, we argue that these attitudes can be changed entirely through unconscious processes. Thus, we conducted a study utilizing the conditioning procedure followed by a stringent test of contingency awareness as the only dependent measure.

Method

Participants. Fifty-seven undergraduates participated for course credit. Although the omission of the non-White participants had no effect on the statistical significance of the results, final analyses included only the 55 White participants.

Stimuli. A variety of neutral and unrelated filler words and images comprised the majority of the stimulus items in the conditioning task (e.g., the words *concrete blocks*, *waffles*, and *electrical outlet* and pictures of farming equipment, umbrellas, and a cab driver). Six different items served as the targets, one per block. These all consisted of pictures and accompanying verbal descriptions of people playing various roles (e.g., a picture of a woman jogging, the verbal descriptor "woman jogging").

CSs consisted of 16 photos (8 Black and 8 White) of clearly Black and White individuals in different occupational roles. The Black and White photo sets were matched in terms of the status and independence of the occupations depicted, as determined by previous research (e.g., a Black male potter and a White male bricklayer) (Fazio & Dunton, 1997).

USs were positive and negative words and images. Words were chosen whose connotations were as purely

evaluative as possible. Images were chosen for their clarity, simplicity, and valence. In every case, positive USs were always evaluated more positively than negative USs according to either our own data or data provided by Ito, Cacioppo, and Lang (1998). A full list of CSs and USs is provided in the appendix.

Procedure. Participants entered the lab under the belief that they would be participating in two unrelated studies. First, they were introduced to the conditioning procedure as a study about attention and rapid responding, where they would be vigilant for specific target items that would periodically appear. Over the course of six blocks, they viewed several hundred images, sometimes alone and sometimes in pairs. Most images, including targets and CS-US pairs, appeared in the center of the screen if alone, and each image of a pair appeared in the center of the left and right halves. Some filler images appeared in randomly selected locations, and some appeared in pairs. For each block, a prespecified target item appeared at random 10 times. To ensure that participants attended to both the words and the images, they were instructed to press a button whenever they saw an image of the target or a verbal descriptor of the target. Sheets with images and descriptors of the targets were provided. For a given block, 8 CS-US pairs (4 Black and 4 White) were presented, totaling 24 pairings for each race across the six blocks. Pairs were constructed from the CS photos and 48 US words and images and were presented in a simultaneous configuration. Each US was presented only once throughout the conditioning procedure, but each CS was presented 3 times. Each CS-US pair was preceded and followed by a blank screen. Sixteen additional blank screens, randomly dispersed throughout each block, were included to reduce any sense of rhythm in the presentation. To obfuscate the CS-US pairs, other filler items consisting of people performing various roles also were presented, sometimes with neutral words and images. Some filler pairs were presented between 5 and 15 times throughout the course of the procedure so as to obscure the repetition of the CSs. Image duration was set to 1.5 s (0-s intertrial interval [ITI]). Participants were encouraged to pay attention throughout the conditioning procedure and to continually focus on the screen to facilitate rapid responding. They were able to rest briefly between blocks when they were shown the target for the following block.

A control condition was constructed to equate exposure to all stimuli without systematically pairing CSs and USs, thus controlling for effects of familiarity or mood. Instead of being paired, CSs and USs were presented separately and randomly. Because of this CS-US pair separation, each block in the control condition had eight additional trials (which lengthened each block by a mere 12 s). Thus, the difference between the two conditions was whether the CS and US appeared on the screen simultaneously or independently. Participants were randomly assigned to either the experimental or control condition.

Dependent measure. A contingency estimation task, modeled after that used by Olson and Fazio (2001), immediately followed the conditioning task. Participants were told that some of the images they saw in the surveillance task may have been presented together. They then viewed several pairs of images, presented sequentially, and were asked to estimate whether the two images appeared together on a –2 (*I'm confident that the 2 items never appeared together*) to a 0 (*don't know*) to a +2 (*I'm confident that the 2 items appeared together at least once*) scale. Thirty-two critical trials consisted of 4 of each randomly chosen Black and White CSs presented with 2 positive USs and 2 negative USs each. Sixteen filler pairs that were and were not actually seen during the conditioning phase also were presented. Presentation order was randomized. They were then debriefed, thanked, and dismissed.

Results

Data preparation. Responses to the items from the contingency estimation task were averaged for each race and valence combination, resulting in four means for each participant (Black-positive, Black-negative, White-positive, and White-negative). Each score reflects the extent to which the participant believed a photo of a given racial category was presented simultaneously with an image of a given valence.

Main analyses. Contingency estimates were submitted to a 2 (condition: experimental vs. control) × 2 (race: Black vs. White) × 2 (valence: positive vs. negative) mixed ANOVA, with repeated measures on the latter two factors. If participants did become aware of our critical pairings of Blacks with positively valenced items and Whites with negatively valenced items, we would expect to see an interaction involving condition, CS race, and US valence. Specifically, participants in the experimental condition would be expected to report higher contingency estimates of Black-positive and White-negative pairs, whereas control participants' estimates involving positive and negative USs should not differ between the races. Instead, analyses revealed only a marginal and difficult-to-interpret Condition × Valence interaction, $F(1, 54) = 3.76$, $p = .06$. Participants were slightly more likely to believe that photos of both races were associated with negative ($M = .76$, $SD = .46$) rather than positive ($M = .64$, $SD = .51$) items in the experimental condition, $t(27) = 1.38$, $p = .18$, and were slightly more likely to believe both races were associated with positive ($M = .74$, $SD = .58$) rather than negative ($M = .60$ $SD = .72$) items in the control condition, $t(27) = 1.39$, $p = .18$. Of importance, race

neither qualified this interaction, $F(1, 54) = 1.58$, $p > .20$, nor produced any other significant or marginal effects (all ts < 1.5, all ps > .20).

Discussion

Experiment 1 provided evidence that participants were unable to recall any of the CS-US contingencies. Experimental participants, who viewed repeated pairings of race and valence, were unable to accurately recognize these image pairs. More specifically, they were no more likely to correctly report that Blacks were paired with positive items (and Whites with negative items) than the reverse. Moreover, their estimates regarding which images appeared with exemplars of the two races did not differ from control participants, for whom there was no association between race and valence. Hence, participants displayed no explicit recollection of the pairings they saw.

However, a skeptic could claim that some participants might still have developed an intuition that Blacks were associated with positive (or Whites with negative) items. That is, participants might have become aware of the covariation between race and valence without being able to identify actual examples of these pairings. We have no reason to believe that this less-episodic form of awareness is any less legitimate than explicit memory of the actual pairings, and our goal is to demonstrate as rigorously as possible that participants were not aware—in any way—of the CS-US pairings. Thus, Experiments 2 and 3 ended with an open-ended measure of awareness that prompted participants to report their intuitions about images that may have appeared together in the conditioning procedure, including a specific item addressing race and valence. If participants still display no awareness of the pairings using this two-pronged approach, we can conclude with more confidence that any attitude change brought about by the conditioning procedure occurs implicitly.

EXPERIMENT 2: ATTITUDE CHANGE

Experiment 2 sought to address the effectiveness of the conditioning procedure on actual attitude change. An identical conditioning procedure was employed, followed by a priming measure of automatically activated racial attitudes. Participants who received Black-positive and White-negative pairings were predicted to show more positive attitudes toward Blacks than control participants, who observed the same stimuli but without any pairings between race and valence. The priming measure was employed because it (a) is not susceptible to the contamination of social desirability biases that affect explicit measures, (b) as reviewed earlier, has proved to reliably predict important race-related race behavior in many previous studies, and (c) does not prompt participants to categorize the primes in any particular way (Olson & Fazio, 2003), which helps to rule out the possibility that any observed attitude change is due to a temporary recategorization of the primes.

Participants. One hundred and nine White, female undergraduates participated for course credit. Of these, 9 were excluded for committing a high number of errors (> 20%) on the priming task and 2 were excluded because they reported some form of awareness of the conditioning procedure on the postexperimental questionnaire.[1]

Materials and procedure. Participants were recruited under the belief that they would be participating in two separate experiments. The first was the conditioning procedure, which was identical to that described in Experiment 1. The second presumed experiment involved the dependent measures, which are described below.

Dependent measures. The principle measure consisted of Fazio et al.'s (1995) priming measure of racial attitudes. The procedure involved five phases. The first was designed to obtain baseline response time data. Participants were told to respond to positive and negative adjectives appearing on the screen by pressing a corresponding button. Each trial began with the presentation of a string of asterisks followed by a randomly selected target adjective that remained on the screen until the participant responded, or for a maximum of 1.75 s; 2.5 s separated each trial. Participants were instructed to respond as quickly and accurately as possible. There were two blocks of 24 trials each, and a practice block preceded the critical blocks.

Phases 2 and 3 were designed to prepare participants for the priming phase and obscure our interest in race. In phase 2, participants were instructed to simply attend to a number of faces (16 color yearbook photos of Black, White, and Asian faces). Each photograph was presented twice—once for each of two blocks—and was presented for .5 seconds. Phase 3 involved a recognition test of the faces presented in phase 2.

Participants were told that phase 4 (the actual priming phase) involved combining phases 1 and 2, and that if word meaning identification was truly an automatic skill, the addition of a face learning task should not inhibit performance on the adjective connotation task. Face primes consisted of 48 photos of Blacks, Whites, Asians, and Latinos. These photos, unlike those employed during the conditioning procedure, were facial images alone and were devoid of any occupation conveying information. On a given trial, a face-prime was presented for 315 ms, followed by a 135-ms interval, followed by the target adjective (the same 24 adjectives from phase 1). Throughout four blocks, each face prime was presented with two positive and two negative adjec-

tives. Thirty-two of the 48 trials per block consisted of Black-White pairs of a given gender, where the same adjectives followed each photo of the pair. Additional filler trials with White, Asian, and Latino faces were included to obscure the experiment's focus on race. Phase 5 consisted of the face recognition task that participants had been led to expect and was similar to phase 3.

After completing the priming measure, participants were asked to complete three unrelated questionnaires that were supposedly being pilot-tested. In fact, these were explicit measures included for exploratory purposes. Participants completed Dunton and Fazio's Motivation to Control Prejudiced Reactions Scale (MCPR) and McConahay's (1986) Modern Racism Scale (MRS), followed by a feeling thermometer measure of attitudes toward various groups. For the latter measure, participants were given a list of 16 different groups (e.g., Blacks, nurses, lawyers, gay men, and lesbians) and were asked to assign a value from 0 (*extremely unfavorable*) to 100 (*extremely favorable*) to each group according to how coolly or warmly they felt toward it. Participants were assured confidentiality and they completed the measures in private, individual cubicles.

After completing the explicit measures, they were asked to complete a postexperimental questionnaire, which was actually the open-ended measure of awareness of the CS-US pairings. This measure was designed to complement the recognition measure of awareness from Experiment 1 by providing participants with an opportunity to communicate any intuition about the covariation between race and valence without any accompanying memory of a specific CS-US pair. The questionnaire began with the question, "In the first session (where you were playing the role of the security guard), did you notice anything unusual about the way the words and pictures were presented?" and each question thereafter was designed to provided incremental "hints" as to the true nature of the experiment (e.g., "Did you notice anything unusual about the words and pictures that were presented with pictures of people of different races? Which ones?"). They were instructed to answer each question as honestly as possible but not to go back to a question once they had answered it. Participants were then debriefed, thanked, and dismissed.

Results

Awareness. No participant reported any amount of awareness of the CS-US pairings on the postexperimental questionnaire prior to the item that explicitly inquired as to the items paired with Blacks and Whites. At that point, 1 participant in the experimental condition correctly responded that Blacks had been paired with positive items and Whites with negative items, and 1 participant in the control condition correctly identified the experimental hypothesis. Neither participant was included in subsequent analyses.

Primary analyses. Our principal hypothesis was that participants in the experimental condition would appear more positive toward Blacks than those in the control condition on our key dependent measure, the priming measure of attitudes. Baseline latencies were computed for each participant for each adjective by averaging the latencies of the two presentations of each adjective in phase 1. The latency for each adjective's presentation during the priming phase (phase 4) was then subtracted from this score, yielding a facilitation score for each face-adjective combination.[2] Mean facilitation scores on the two positive and two negative adjectives were then computed for each of the 32 critical faces for each participant. Means of same-race and same-valence facilitation scores were computed for each participant, and analyses are based on these variables.

A 2 (race of prime: White vs. Black) × 2 (target valence: positive vs. negative) × 2 (condition: control vs. experimental) mixed-factor ANOVA was performed on participants' facilitation scores from the priming task, with repeated measures on the former two factors. This analysis revealed the predicted three-way interaction, $F(1, 97) = 4.78$, $p < .05$. In the control condition, the two-way Race of Prime × Target Valence interaction indicative of prejudice was found, $F(1, 45) = 3.27$, $p = .07$ (see Figure 1A). Controlling for baseline response latencies to the target adjectives alone, participants were quicker to respond to positive words following White primes ($M = -13.2$, $SD = 84.6$) compared to Black primes ($M = -27.0$, $SD = 82.3$), $t(45) = 2.53$, $p < .01$, indicating a prejudice in favor of Whites. No difference was found in response latencies to negative words following White primes ($M = -17.8$, $SD = 76.4$) compared to Black primes ($M = -19.4$, $SD = 77.7$, $t < 1$). In the experimental condition, this interaction was not significant, $F(1, 52) = 1.77$, $p = .19$ (see Figure 1B). Moreover, the pattern of response latencies suggests a positive bias in favor of Blacks. That is, participants were slightly quicker to respond to positive words following Black primes ($M = -18.1$, $SD = 56.4$) compared to White primes ($M = -25.2$, $SD = 55.8$), although not significantly so, $t(52) = 1.42$, $p = .16$. Response latencies to negative words in the experimental condition also did not differ significantly depending on whether they were preceded by Black primes ($M = -26.6$, $SD = 62.6$) versus White primes ($M = -24.2$, $SD = 69.0$).

As Table 1 illustrates, the explicit measures of racial attitudes and motivation to control prejudiced reactions revealed no differences between conditions (all ts < 1.5, ps > .18).

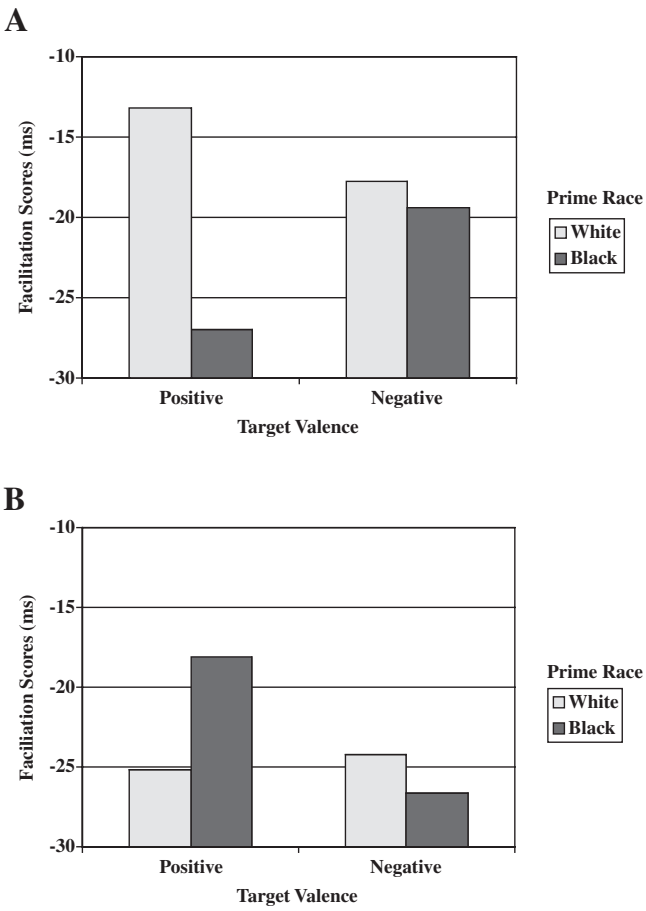

Figure 1 Race of Prime × Valence of Target interaction for (A) the control condition and (B) the experimental condition, Experiment 2.
NOTE: Higher numbers reflect greater facilitation.

Discussion

To summarize, the conditioning procedure was successful. Participants in the experimental condition showed less negative, automatically activated racial attitudes than control participants, as indicated by the priming measure. Of importance, this attitude change was observed even though participants in the two conditions were exposed to identical images during the conditioning procedure and even though the photos used in the priming procedure differed from those that appeared in the conditioning phase. Participants' attitudes toward new exemplars of the category "Blacks" were affected by the conditioning procedure.

Finally, the open-ended measure of awareness complemented the recognition measure from Experiment 1 in demonstrating participants' lack of awareness of the pairings. Only one experimental participant (whose data were not included in analyses) displayed any awareness of the association between race and valence. Thus, it appears that the cover story and procedure were effective at minimizing both explicit recognition of specific US-CS pairs and the development of intuitions about the connection between race and valence in the procedure.

TABLE 1: *Ms* (*SD*s) for Each Explicit Measure by Condition (Experiment 2)

Measure	*Experimental*	*Control*	t
Concern factor of MCPR[a]	.18 (.98)	.17 (.91)	<1
Restraint factor of MCPR	–.18 (1.05)	.01 (.97)	<1
MRS	–.81 (.74)	–1.02 (.70)	1.42
Raw feeling thermometer	76.00 (18.4)	76.20 (19.7)	<1
Standardized feeling thermometer	.42 (.58)	.41 (.69)	<1

NOTE: MCPR = Motivation to Control Prejudiced Reactions Scale; MRS = Modern Racism Scale.
a. Factor analyses replicated the two-factor structure of the MCPR (Dunton & Fazio, 1997). Factor scores were computed for each participant based on varimax rotation of the factor solution.

EXPERIMENT 3: PERSISTENCE

Participants completed the same EC procedure in Experiment 3 but completed the dependent measure only after a 2-day delay to investigate whether the conditioning procedure would have any lasting effect.

Participants. Fifty-eight White, female undergraduates participated for course credit. Six participants failed to return for the second session and 5 participants were excluded for committing too many errors (> 20%) on the priming measure.

Materials and procedure. Both the conditioning procedure and the dependent measures were identical to those used in Experiment 1. Also, as in Experiment 1, participants were led to believe that they would be partaking in two separate experiments but that they would have to return at the same time in 2 days to complete the second experiment. No participant expressed suspicion about the potential connection between the two sessions. After completing the conditioning phase in session one, participants signed up for the second session, at which time they completed the dependent measures. Participants were then debriefed, thanked, and dismissed.

Results and Discussion

Awareness. No participant correctly identified the CS-US contingencies on the postexperimental questionnaire, and no one identified the hypothesis of the experiment.

Primary analyses. Data preparation proceeded just as described in Experiment 2. Primary analyses involved comparing participants from the two conditions on the various measures of racial attitudes. Participants' facilitation scores from the priming measure were subjected to

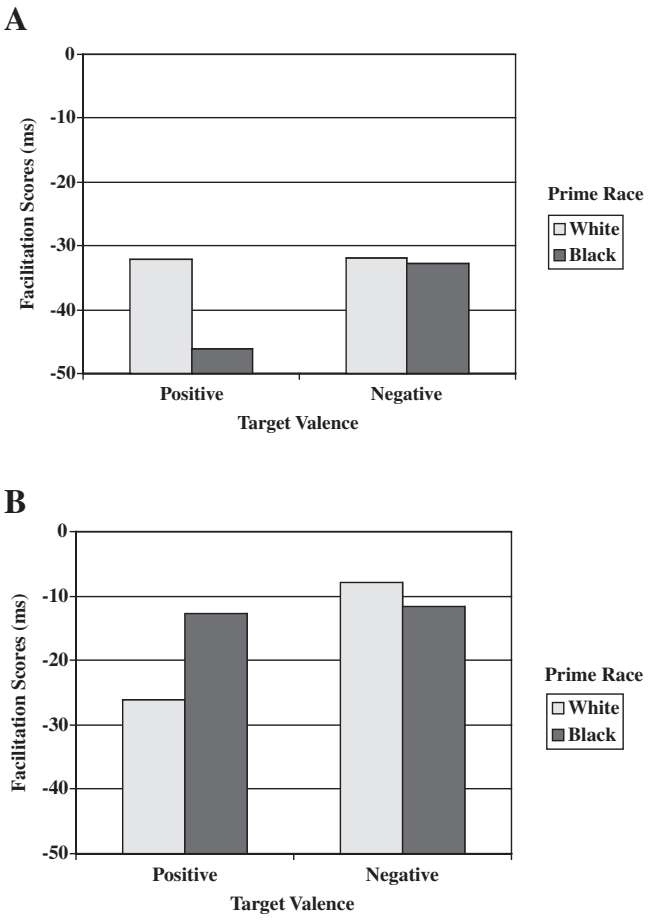

Figure 2 Race of Prime × Valence of Target interaction for (A) the control condition and (B) the experimental condition, Experiment 3.

NOTE: Higher numbers reflect greater facilitation.

a 2 (race of prime: White vs. Black) × 2 (target valence: positive vs. negative) × 2 (condition: control vs. experimental) mixed-factor ANOVA, with repeated measures on the former two factors. The predicted three-way interaction emerged, $F(1, 45) = 4.50$, $p < .05$. The two-way Race of Prime × Target Valence interaction was not significant in the control condition, indicating that, on average, participants were not prejudiced against Blacks, $F(1, 25) = 1.6$, $p > .20$ (see Figure 2A).[3] Participants in the control condition responded to negative targets with roughly equal speed following White primes ($M = –32.0$, $SD = 58.5$) versus Black primes ($M = –32.8$, $SD = 77.2$, $t < 1$) but they were slightly quicker to identify positive targets following White primes ($M = –32.2$, $SD = 57.5$) versus Black primes ($M = –46.1$, $SD = 74.1$), suggesting a slight bias in favor of Whites, $t(25) = 1.78$, $p = .09$. However, the experimental condition revealed a Race of Prime × Target Valence interaction indicative of prejudice in favor of Blacks, $F(1, 20) = 3.44$, $p = .07$ (see Figure 2B). Here, participants responded to negative targets following White primes ($M = –8.0$, $SD = 111.4$) and Black primes ($M = –11.7$, $SD = 97.7$) with equivalent speed ($t < 1$), but they were somewhat quicker to respond to positive targets following Black primes ($M = –12.8$, $SD = 77.6$) relative to White primes ($M = –26.2$, $SD = 71.4$), $t(20) = 1.71$, $p = .10$.

TABLE 2: *Ms* (*SD*s) for Each Explicit Measure by Condition (Experiment 3)

Measure	*Experimental*	*Control*	t
Concern factor of MCPR[a]	.43 (.66)	.15 (.82)	1.25
Restraint factor of MCPR	–.98 (.87)	–.73 (.92)	<1
MRS	–.63 (.69)	–.82 (.56)	1.05
Raw feeling thermometer	66.91 (18.5)	75.42 (22.1)	1.43
Standardized feeling thermometer	–.08 (.75)	.34 (.89)	1.73

NOTE: MCPR = Motivation to Control Prejudiced Reactions Scale; MRS = Modern Racism Scale.

a. Given the smaller sample size of Experiment 3, MCPR factor scores were based on a factor analysis that also included a sample of 415 mass-survey participants from the same population who completed the scale during the same time period. Once again, the two-factor structure observed by Dunton and Fazio (1997) was replicated.

The explicit measures did not vary between the two conditions (*t*s < 1), with the exception of the standardized feeling thermometer measure of attitudes toward Blacks, which displayed an inexplicable trend toward greater negativity toward Blacks in the experimental relative to the control condition, $t(46) = 1.73$, $p = .09$. Table 2 provides means and comparisons by condition for all measures.

In sum, predictions were again confirmed—experimental participants showed more positive racial attitudes on the priming measure than did control participants, even with a 2-day delay between the conditioning and attitude measurement phase. Moreover, none of the participants reported any intuitions about the covariation between race and valence.

GENERAL DISCUSSION

In the experiments reported here, Whites' automatically activated racial attitudes toward Blacks were made more positive using an EC procedure. Specifically, through the inclusion of pairings of Blacks with positive words and images and Whites with negative words and images in an unrelated detection task, a well-validated priming measure of racial attitudes revealed relatively less automatically activated negativity in response to Blacks (Experiments 2 and 3). Moreover, as demonstrated in Experiment 3, this attitude change persisted for 2 days.

We believe the change found in experimental participants' evaluations of Blacks to represent change in "the judgment of the object," not change in the "object of judgment," to return to the quotation from Asch (1940,

p. 458). As mentioned earlier, a number of recent studies have demonstrated the malleability of implicit measures of attitudes (Blair, 2002). Many of these studies have shown that prompting participants to recategorize members of outgroups as members of liked groups leads to the activation of a positive evaluation (e.g., Dasgupta & Greenwald, 2001; Sinclair & Kunda, 1999; Wittenbrink et al., 2001). Prejudicial responses are averted in these situations by either prompting participants to avoid the kinds of categorization processes that lead to activation of prejudice or by altering the response label construals that govern performance on the implicit measure. Presumably, memory structures that normally lead to the activation of prejudiced responses remain intact. In the experiments reported here, experimental participants were neither instructed nor motivated to categorize Black targets as members of a specific category. In fact, our interest in race was obscured throughout the entire experimental process (until, perhaps, the explicit measures). Because we can rule out these other means of temporarily altering responses on implicit measures, we can conclude that it was participants' associations in memory—those that prompt prejudicial responses across time and contexts—that must have changed.

However, this change was not observed on explicit measures, and it is important to consider why. Explicit measures are known to be affected by social desirability biases (e.g., Fazio et al., 1995). Although some individuals will honestly report their racial attitudes on explicit measures, others are motivated to avoid appearing prejudice (Fazio & Dunton, 1997). To the extent that many of the present participants were so motivated, we would not expect explicit measures to reflect the automatically activated attitude directly. Indeed, previous findings suggest that people characterized by both automatically activated negativity toward Blacks and motivation to avoid it are likely to overcorrect for their negativity and appear even more positive than people for whom positivity is activated (Olson & Fazio, 2004). The automatically activated attitudes of participants in the experimental condition grew more positive. Their responses to the explicit measures may have reflected this positivity but proven indistinguishable from the responses of participants in the control condition who corrected for automatically activated negativity. In other words, the two conditions may have ended at the same point of reporting favorability toward Blacks but reached that point through different processes.

Implications for Evaluative Conditioning

Classical conditioning is one of the most basic learning phenomena in existence but its extension to attitudes—evaluative conditioning—has been tenuous (see De Houwer et al., 2001, for an excellent review). Only recently has it been convincingly demonstrated that EC occurs without awareness of the pairings between the attitude object and some other valenced items (e.g., De Houwer et al., 2001; Olson & Fazio, 2001). Little is known about the potential of EC as a tool to change preexisting attitudes. The only related research of which we are aware is that of Cacioppo, Marshall-Goodell, Tassinary, and Petty (1992), who demonstrated that EC has a stronger influence on attitudes toward unknown versus known objects. The present research provides some of the only evidence to date that EC can be used not only to develop attitudes but as a tool for promoting attitude change as well.

One might argue that we have not convincingly demonstrated attitude change because in Experiment 3 there was little evidence of prejudice against Blacks in our control condition. In response to this argument, some samples of Whites may, on average, show a bias against Blacks and other samples may, on average, appear neutral (see Note 2). This does not mean that the former group has racial attitudes and the latter group does not. Individuals' racial attitudes can range from positive to negative, and some Whites have positive automatically activated attitudes toward Blacks. Our point was to show a change in racial attitude in a positive direction compared to an equivalent control group. Moreover, the control participants in Experiment 2 did exhibit negativity toward Blacks on average. Thus, we believe these findings convincingly demonstrate attitude change.

The present research is also unique in that it demonstrates attitude change at the level of the category through conditioning at the level of the exemplar. This is noteworthy due to the nature of the covariation detection it requires of the perceiver. Most prior EC work has utilized relatively impoverished stimuli where the CS-US covariations were one of the few pieces of information actually available, and participants did not have to abstract information about the category from exemplars. Here, participants could construe the stimuli in a number of ways, and race and valence were among a rich array of other potential pieces of information for which covariations might have existed. Given that participants were able to detect covariations between race and valence from such complex stimuli, the possibility that EC occurs in more information-rich, real-world environments seems all the more likely.

To the extent that participants attended to the race-related information in the photos, we would expect them to more effectively encode the covariation between race and valence. Previous research has revealed that individuals for whom race is attitude evoking—those for whom either positivity or negativity is more strongly activated automatically—are more likely to categorize by

race than are those with less extreme reactions. Using a larger set of the same photos we employed as CSs in our experiments, Fazio and Dunton (1997) observed greater attention to race in a later similarity judgment task by individuals with more extreme scores on the priming measure of racial attitudes. The implication is that such people may be more likely to implicitly detect any race-valence covariation.

Finally, the present findings contribute to the EC literature by providing further corroboration of conditioning without awareness of the CS-US pairings. To expand on the meaning of contingency awareness, it is worth noting that two forms are possible. First, someone might recognize specific CS-US pairs without forming the generalization that race and valence are related, and second, someone might develop an intuition that race and valence are related without being able to recall specific examples. Neither form of awareness was evident in the experiments reported here. Specifically, participants displayed no explicit recognition for specific CS-US pairings in Experiment 1, and they expressed no suspicion that race and valence covaried on an open-ended measure of awareness in Experiments 2 and 3. This two-pronged approach to the awareness issue argues against continued claims that EC requires contingency awareness (e.g., Field, 2000).

On the Relative Ease of Reducing Automatic Prejudice

Automatic activation of prejudice is still commonly assumed to be so well-ingrained in the social perceiver's psyche that efforts at prejudice reduction would be best allocated to changing the extent to which perceivers engage in motivated impression formation and social reasoning (Devine & Monteith, 1999). Indeed, it was only after a laborious 480-trial procedure that any change was observed in Kawakami et al.'s (2000) research. Yet, the evidence here is that attitudes changed after only 24 Black-good and White-bad pairings. Most research would suggest that automatic racial prejudice should not be changed with so few trials (as would Kawakami et al.'s research reviewed earlier), so it is critical to consider why the present procedure was so effective.

Without belittling the many differences between the present procedure and Kawakami et al.'s, one important distinction is with respect to participants' awareness of the counterprejudicial associations presented to them. In Kawakami et al.'s work, participants were instructed to say "no" to themselves when they saw a social target followed by a stereotypic trait. That is, they were to take conscious note of stereotypic associations and mentally negate them (and say "yes" to target-trait pairs that were counterstereotypic). Because Kawakami et al. used well-known stereotypes (e.g., skinheads, Blacks), it is likely that at least some participants found themselves recalling their own agreement with the stereotype at the same time they were saying "no." Moreover, they also may have recalled specific stereotype-consistent exemplars, perhaps out of reactance. It is also worth noting that regardless of their level of personal endorsement, participants still saw many outgroup members presented with negative, stereotype-consistent traits. Hence, in terms of what was encoded, both prejudice-consistent and prejudice-inconsistent information were actively presented to participants, and participants may themselves have actively recruited cases of negative stereotype consistency. Perhaps this competition between contradictory associations increased the necessary number of trials before counterprejudicial associations could eventually win out. In the research reported here, participants were neither exposed to any stereotype or prejudice-consistent information nor given an opportunity to recall cases of stereotype consistency. These differences may help explain why so few trials were necessary to reduce prejudice in participants. Of course, these are empirically testable speculations and future research may help to substantiate them.

On a more general level, it is important to note that the Black-positive and White-negative covariations were without exception in the present procedure (i.e., every Black CS was paired with a positive US, and every White CS was paired with a negative US). The real world certainly does not present such perfect covariations, and perhaps of equal importance, it does not present such covariations at the rapid rate employed here. Evidence from neural network models of learning suggests that such condensed or "focused" learning situations may have unique power in changing relatively stubborn, long-term associations in memory (e.g., McClelland, McNaughton, & O'Reilly, 1995). Thus, even well-learned associations may be susceptible to change given the appropriate kinds of input. The rigidity that tends to characterize the human mind, then, may be more a reflection of a stubborn environment than a stubborn psyche. One implication of this analysis is that less than perfect covariations between a given race and a given valence might dramatically weaken the conditioning effect observed here. Research that systematically manipulates the degree of covariation between race and valence in a procedure such as the present one might therefore be informative.

Not only might the degree of covariation between US and CS alter the nature of the attitude change but valence itself may matter as well. In both of the experiments reported here, the effect of the conditioning manipulation was more apparent on the trials involving positive target words than those involving negative targets. The simplest explanation for this difference is that

positive targets in the priming task may be more sensitive to the valence of the primes than may be negative targets. Indeed, the control conditions in each of our experiments displayed greater differentiation by prime on the positive target adjectives than on the negative target adjectives. Earlier experiments also obtained relatively stronger effects on positive targets than on negative targets (e.g., see Fazio et al., 1995; Olson & Fazio, 2002). For this reason, we are hesitant to draw inferences from our having observed more change in response latencies to positive rather than negative targets. This does not necessarily imply that the conditioning procedure only strengthened positive associations with Blacks. It may be equally likely that negative associations with Blacks were weakened as a result of the conditioning. In any case, this issue may have important theoretical and practical implications and will hopefully be informed by future work.

Conclusion

Racial prejudices are often argued to form with little effort or awareness and to manifest with little effort or awareness. Unique to the experiments reported here is that from attitude change to attitude measurement, participants' attention was elsewhere. Indeed, the implicit classical conditioning procedure employed here may mirror the real-world processes that all too often strengthen our existing prejudices. However, the present findings indicate that these same conditioning processes have the potential to ameliorate existing prejudices.

APPENDIX
CS and US Items

Black female CSs: cashier, businessperson, telephone repairperson, nurse
Black male CSs: minister, businessperson, potter, landscaper
White female CSs: police officer, painter, businessperson, pharmacist
White male CSs: professor, trash collector, architect, bricklayer
Negative US words: horrible, annoying, repulsive, appalling, disgusting, sickening, worthless, awful, dreadful, terrible, saddening, upsetting
Negative US images: couple at grave, snowy cemetery, garbage in the sand, dirty dishes, old man near bed of sick woman, crying soldier, auto exhaust, bees, crying boy, cigarette butts, person in contamination suit, trash in street
Positive US words: magnificent, amazing, fabulous, delightful, excellent, outstanding, exciting, fantastic, terrific, awesome, enjoyable, wonderful
Positive US images: hippo in pond, young couple embracing, city skyline, woman holding baby, boy with ice cream cone, Popsicle, puppies, flower bed, man smiling with children, smiling older couple, bunnies with dandelion, baby seal

NOTE: CS = conditioned stimulus; US = unconditioned stimulus.

NOTES

1. The 20% cutoff was a function of a desire to exclude a minimal number but to exclude those who clearly were not engaged in the task. Given the ease of the adjective connotation task, making errors on 20% or more of the trials strongly suggests inattentiveness. Indeed, this error rate was fully 3 standard deviations above the mean. The number of participants omitted because of high errors did not differ by condition for Experiment 2 (4 and 5) or 3 (2 and 3, for the experimental and control conditions, respectively; both chi-squares < 1).

2. The term *facilitation* is used in a relative sense. It is neither surprising nor inconsistent with past research that these facilitation scores are negative given the additional task of face learning during the priming phase (Fazio, Jackson, Dunton, & Williams, 1995).

3. Of the studies using the priming measure of racial attitudes that have been conducted in our lab, the sample has sometimes appeared significantly negative toward Blacks on average (Fazio et al., 1995; Olson & Fazio, 1999; Towles-Schwen & Fazio, 2001) and sometimes not (Fazio & Dunton, 1997; Fazio & Hilden, 2001; Jackson, 1997; Olson & Fazio, 2004; Towles-Schwen & Fazio, in press). Thus, we would speculate that a little more than half of White undergraduates possess negative, automatically activated attitudes toward Blacks and that sampling variability underlies inconsistencies regarding mean differences. What is more important is that the predictive validity of the priming measure has been consistently demonstrated regardless of whether the sample is prejudiced on average.

REFERENCES

Asch, S. E. (1940). Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. *Journal of Social Psychology, 12*, 433-465.
Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.
Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-260.
Blair I., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828-841.
Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28*, 207-233.
Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*, 360-365.
Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800-814.
De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.
Devine, P. G. (1989). Stereotypes and prejudice: Automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-17.
Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 339-360). New York: Guilford.
Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interactions. *Journal of Personality and Social Psychology, 82*, 62-68.
Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510-540.
Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316-326.
Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. New York: Harcourt Brace.

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

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. 75-107). New York: Academic Press.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33*, 451-470.

Fazio, R. H., & Hilden, L. (2001). Emotional reactions to a seemingly prejudiced response. *Personality and Social Psychology Bulletin, 27*, 538-549.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Fazio, R. H., & Olson, M. A. (2003a). Attitudes: Foundations, functions, and consequences. In M. Hogg & J. Cooper (Eds.), *The Sage handbook of social psychology* (pp. 139-160). London: Sage.

Fazio, R. H., & Olson, M. A. (2003b). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 97- 116). New York: Guilford.

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*, 13-36.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231-254). New York: Guilford.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology, 40,* 357-365.

Greenwald, A. G., McGhee, D., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association task. *Journal of Personality and Social Psychology, 74*, 1469-1480.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin, 24*, 855-879.

Jackson, J. R. (1997). *Automatically activated racial attitudes.* Unpublished doctoral dissertation, Indiana University.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training on the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871-888.

Krosnick, J. A., Betz, A. L., Jussim, L. J., Lynn, A. R., & Stephens, L. (1992). Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin, 18*, 152-162.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5-18.

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

Macrae, C. N., Bodenhausen, G. S., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*, 397-407.

Macrae, C. N., Bodenhausen, G. S., Milne, A. B., Thorn, T. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33*, 471-489.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

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.

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

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

Olson, M. A., & Fazio, R. H. (1999, May). *Nonverbal leakage of automatically activated racial attitudes.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, Illinois.

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

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20*, 89-104.

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

Olson, M. A., & Fazio, R. H. (2004). Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology, 26*,1-12.

Razran, G. H. S. (1938). Conditioning away social bias by the luncheon technique. *Psychological Bulletin, 37*, 481.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice stereotypes. *Journal of Personality and Social Psychology, 81*, 856-868.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367-447.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885-904.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology, 71*, 888-898.

Towles-Schwen, T., & Fazio, R. H. (2001). On the origins of racial attitudes: Correlates of childhood experiences. *Personality and Social Psychology Bulletin, 27*, 162-175.

Towles-Schwen, T., & Fazio, R. H. (2002, February). *Interracial roommate relationships: The role of automatically-activated racial attitudes.* Paper presented at the annual meeting of the Society for Personality and Social Psychology, Savannah, Georgia.

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

Towles-Schwen, T., & Fazio, R. H. (in press). Automatically activated racial attitudes as predictors of the success of interracial roommate relationships. *Journal of Experimental Social Psychology.*

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

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

Received November 12, 2004

Revision accepted August 23, 2005