Journal of Personality and Social Psychology
2001, Vol. 81, No. 5, 856–868

0022-3514/01/$5.00 DOI: 10.1037//0022-3514.81.5.856

"Unlearning" Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes

Laurie A. Rudman, Richard D. Ashmore, and Melvin L. Gary
Rutgers, The State University of New Jersey

The present research suggests that automatic and controlled intergroup biases can be modified through diversity education. In 2 experiments, students enrolled in a prejudice and conflict seminar showed significantly reduced implicit and explicit anti-Black biases, compared with control students. The authors explored correlates of prejudice and stereotype reduction. In each experiment, seminar students' implicit and explicit change scores positively covaried with factors suggestive of affective and cognitive processes, respectively. The findings show the malleability of implicit prejudice and stereotypes and suggest that these may effectively be changed through affective processes.

> I refuse to accept the view that mankind is so tragically bound to the starless midnight of racism and war that the bright daybreak of peace and brotherhood can never become a reality.
>
> —Martin Luther King, Jr.

Ten years after the landmark Supreme Court ruling that legislated equal opportunity for African Americans in education, Martin Luther King, Jr. accepted the Nobel Peace Prize. As King's quote illustrates, his remarks were characterized by optimism and resolve. Having weathered the storms of desegregation and the violent backlash against it (Jones, 1972), he accepted this honor on behalf of the Civil Rights Movement, even while acknowledging that the movement had not yet fulfilled its promise to African Americans.

Nearly 40 years later, that promise remains to be fulfilled. Despite legislation and policies (e.g., affirmative action) designed to redress a history of oppression, Blacks continue to suffer discrimination in the areas of employment, housing, and health care. Blacks do not enjoy the same justice system, police protection, or voting rights as do Whites. Although they putatively have equal access to education, the quality of their education is not equal. So what, exactly, has changed?

It is now illegal, as well as immoral, to discriminate against people on the basis of group membership. Enforced compliance with the Civil Rights Act has led to a dramatic decrease in the overt expression of racism (Schuman, Steeh, Bobo, & Krysan, 1997) and a commensurate increase in normative pressures to be nonprejudiced (Dunton & Fazio, 1997; Plant & Devine, 1998). Indeed, if researchers were to rely solely on self-report measures of attitudes toward Blacks, they would be hard-pressed to conclude anything other than that prejudice has become, if not outdated, at least unfashionable. In reality, however, prejudice continues to dog Americans' footsteps, even as we make progress toward an egalitarian ideal (Eberhardt & Fiske, 1998).

This reality comes starkly into focus when researchers examine people's actions, rather than their attitudes, toward Blacks (Crosby, Bromley, & Saxe, 1980), including when the behavior can be justified by the selective interpretation of ambiguous information (Dovidio & Gaertner, 2000) or expressed in covert ways (Beal, O'Neal, Ong, & Ruscher, 2000). It is also disheartening that children, who may be less cognizant of egalitarian norms than adults, continue to report prejudiced attitudes (see Bigler, 1999, for a review). Furthermore, when adults' attitudes are measured using more subtle instruments (e.g., McConahay, 1986; Sears, 1988), or bogus pipeline techniques (Roese & Jamieson, 1993), racism is often exposed. Finally, when attitudes are measured when using techniques that do not rely on respondents' willingness or ability to report their opinions, pervasive anti-Black biases are often revealed (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995). For example, people who report feeling "exactly the same" about Whites and Blacks nonetheless show evidence of implicit preference for Whites (Greenwald, McGhee, & Schwartz, 1998).

In summary, despite dramatic reductions in self-reported prejudice, other indicators suggest that racism persists, even on the part of avowed egalitarians. Theoretically, there are at least three reasons why this might be so. First, Whites may repress anti-Black biases because they conflict with an egalitarian self-image (Gaertner & Dovidio, 1986). Second, even people aware of their prejudices may be unwilling to express them because of normative pressures (Dovidio & Fazio, 1992; Fazio et al, 1995; Plant & Devine, 1998). Third, self-report measures are, by definition, subject to respondents' conscious introspection. What they cannot show is the extent to which people have acquired automatic (or overlearned) associations that reflect negatively on Blacks (i.e., implicit prejudice and stereotypes). These associations stem from socialization processes that may not be directly taught, but which nonetheless accumulate as a result of living in a culture that has traditionally favored some groups over others (Devine, 1989). If

Laurie A. Rudman, Richard D. Ashmore, and Melvin L. Gary, Department of Psychology, Rutgers, The State University of New Jersey.

This research was partially supported by National Science Foundation Grant SBE-9807970.

Correspondence concerning this article should be addressed to Laurie A. Rudman, Department of Psychology, Tillett Hall, Rutgers, The State University of New Jersey, 53 Avenue E, Piscataway, New Jersey 08854-8040. Electronic mail may be sent to rudman@rci.rutgers.edu, ashmore@rci.rutgers.edu, or mgary@rci.rutgers.edu.

people do not know they possess implicit biases, they cannot accurately report them.

Response Latency Measures

The principle that intergroup biases may be inaccessible is central to the social cognitive approach to orientation assessment (Greenwald & Banaji, 1995). In this approach, explicit orientations consist of attitudes and beliefs that people are willing and able to report. By contrast, implicit orientations consist of automatic associations (e.g., between Blacks and criminality) that are unavailable to introspection; therefore, only implicit measures can detect them.[1] Of these, response latency measures have shown great promise vis-à-vis assessing attitudes and beliefs. The techniques include semantic and evaluative priming tasks (e.g., Dovidio et al., 1997; Fazio et al., 1995; Kawakami, Dion, & Dovidio, 1998; Wittenbrink, Judd, & Park, 1997), and associative categorization tasks (e.g., Greenwald et al., 1998), but results have been similar. For the most part, these investigations have shown pervasive implicit anti-Black orientations (on the part of Whites) that are modestly related to conscious opinions (for a meta-analysis, see Dovidio, Kawakami, & Beach, 2001). These findings support conceptualizing implicit and explicit orientations as related but distinct concepts (Devine, 1989; Wilson, Lindsey, & Schooler, 2000).

The absence of strong convergence with explicit attitude and stereotype measures has led researchers to validate implicit measures through other means. These include showing known groups validity (e.g., Fazio et al., 1995; Greenwald et al., 1998), convergent validity among different implicit measures (Brauer, Wasel, & Niedenthal, 2000; Cunningham, Preacher, & Banaji, 2001; Rudman & Kilianski, 2000), and predictive utility. For example, friendly behavior toward Blacks is negatively predicted by implicit prejudice, whether assessed by semantic priming (Dovidio et al., 1997), evaluative priming (Fazio et al., 1995), or associative categorization (McConnell & Leibold, 2001). In addition, discriminating against female job applicants is predicted by implicit stereotypes, whether assessed by semantic priming (Rudman & Borgida, 1995) or associative categorization (Rudman & Glick, in press). Finally, implicit anti-Black prejudice has been shown to covary with amygdala activation in Whites exposed to photos of Blacks (Phelps et al., 2000). Because the amygdala is associated with emotional learning, including fear conditioning, these findings suggest that implicit biases are linked to perceptions of anxiety or threat (see also Amodio, Harmon-Jones, & Devine, 2000). Taken together, the research warrants conceptualizing implicit measures as indicators of individual differences in the propensity to automatically evaluate social groups unfavorably.

This is not meant to imply that automatic prejudice is universal or inevitable (cf. Bargh, 1999; Crosby et al., 1980). For example, Devine and her colleagues found that people who reported high internal (and low external) motives to be nonprejudiced showed relatively low levels of implicit prejudice, using both response latency (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2000) and physiological (Amodio et al., 2000) measures. Thus, there appear to be people for whom prejudicial responses are less likely to be automatic, specifically, those who have internalized egalitarian norms (see also Moskowitz, Gollwitzer, Wasel, & Schaal, 1999).

Intervention Strategies

Evidence for the persistence of anti-Black feelings and beliefs has spurred a variety of intervention efforts, most of which target explicit racism. For example, schoolchildren routinely undergo diversity training, often through immersion in multicultural curricula (Bigler, 1999). When the effectiveness of this training has been examined, results have typically been disappointing (Banks, 1995). Indeed, children in experimental conditions have often shown more racial bias after the intervention, compared with control students (Bigler, 1999). Similarly, Hewstone (1996) critically reviewed the literature on adult interventions, many of which rely on the contact hypothesis (Allport, 1954). One problem with these interventions is that they seldom yield results that generalize beyond the specific contact situation to group-based attitudes as a whole (most likely because of subtyping effects; Weber & Crocker, 1983). Interventions that rely on "color-blind" strategies, in which people are encouraged to suppress their category-based stereotypes in favor of more personalized judgments, appear to be particularly ineffective (Wolsko, Park, Judd, & Wittenbrink, 2000) and may even backfire (Schofield, 1986). Because suppression may lead to rebound (Wegner, 1994), stereotypes can, ironically, become more accessible as a result of color-blind interventions. Interventions based on making group membership salient (i.e., those that stress appreciation, rather than the elimination, of group differences) have been somewhat more successful (e.g., Johnston & Hewstone, 1992; Wilder, 1984; Wolsko et al., 2000). However, even these are subject to limitations, including the tendency for greater negative, as well as positive, generalized change to occur (Hewstone, 1996).

A potentially more serious problem is that people are often forced to undergo diversity training. In this case, the message may result in backlash because of reactance (i.e., the need to preserve psychological autonomy; Brehm, 1966). Under these conditions, people may perceive a threat to their freedom of expression or be offended by the implication that they are prejudiced. The impact of enforced multicultural training is not a trivial one, as many education administrators have embraced it as a catalyst for overcoming discrimination. For example, a recent survey found that 81% of U.S. colleges and universities have used diversity workshops, yet none of these institutions have undertaken an evaluation of their effect (McCauley, Wright, & Harris, 2000). Aware of the potential problems of enforced "political correctness," Plant and Devine (2000) recently examined people's reactions to being pressured to comply with a pro-Black request. In three experiments, they found that some people (specifically, those high in external, but low in internal, motives to be nonprejudiced) responded with anger and, in turn, lashed out at African Americans on attitude and behavior measures. At the very least, their findings suggest that pressure to conform to pro-Black standards does not have uniformly positive effects.

[1] The terms *implicit* and *explicit* are not meant to imply a literal or exhaustive dichotomy. Rather, automatic processes often involve some component of awareness, just as controlled processes may be routinized to varying degrees (Bargh, 1989; see also Shiffrin, 1988; Wegner & Bargh, 1998, for discussions of the interactivity of automatic and controlled processes).

This review of diversity training considers the limited, or even negative, effects of such training on nonvolunteers (i.e., students in classrooms or research subjects forced to comply with external demands). What about people, though, who volunteer for multicultural education? By definition, volunteers have chosen to learn about prejudice and should, therefore, be less likely to show reactance. Many U.S. college campuses offer courses aimed at cultural pluralism, including those designed to reduce anti-Black biases. Students who voluntarily enroll in these courses are presumably "ready and willing" to respond to the curriculum's message. As a result, their educational experience should have primarily positive effects. Specifically, they should show reduced prejudice and stereotyping after the course is over, compared with the beginning.

Surprisingly, the impact of multicultural training on volunteers has, to date, not been investigated. Educators appear to assume this type of coursework is beneficial for students without examining its effectiveness. Thus, a major goal of the present research was to examine the effect of diversity education on student volunteers. In view of the prevalence of automatic anti-Black biases, we assessed the potential impact of multicultural training on students' implicit, as well as explicit, prejudice and stereotypes. If diversity education is shown to reduce either students' self-reported or automatic biases, then its putative benefits will receive support—at least when students select themselves as intervention candidates.

Are Implicit Biases Permanent?

Because volunteers for diversity training are "ready and willing" to undergo transformation, we expected students enrolled in a multicultural course to show reduced prejudice and stereotypes, at least at the level of self-reports. The question of whether they might show reduced implicit biases was more speculative. Implicit orientations are conceptualized to stem, at least in part, from long-standing status differences between groups. As a result, researchers have reasonably assumed that they are stable, enduring, and resistant to change (e.g., Bargh, 1999; Dovidio et al., 1997; Fazio et al., 1995; Greenwald et al., 1998). Nonetheless, if implicit orientations have their basis in overlearned associations, then they should be amenable to change (i.e., unlearning; Devine, 1989; Monteith, 1993).

To date, laboratory efforts to change implicit biases have supported this view. For example, subjects who extensively practiced reversed stereotypic associations showed decreased automatic stereotype activation (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). Moreover, there is growing evidence that implicit associations are sensitive to environmental influences, including priming effects (Dijksterhuis & van Knippenberg, 1996; Karpinski & Hilton, 2001; Rudman & Borgida, 1995). For example, Dasgupta and Greenwald (2001) exposed people to positive Black exemplars (e.g., Denzel Washington) and negative White exemplars (e.g., Timothy McVeigh). This manipulation effectively reduced automatic anti-Black evaluation, both immediately and 24 hr later. Similarly, subjects instructed to imagine strong (vs. dainty) women showed reduced implicit gender stereotyping (Blair, Ma, & Lenton, 2001). Taken together, the results suggest that implicit associations can be modified, at least temporarily, by focusing subjects' attention on subtypes of group members or by activating links in the cognitive network that are antithetical to traditional stereotypes (Bodenhausen & McCrae, 1998; Kunda & Thagard, 1996). What is less clear is whether implicit biases might be moderated through real-world experiences—including diversity education—and, if so, what factors might covary with these changes. This was the focus of the present research.

Research Objectives

In two experiments, we examined whether education in prejudice and intergroup conflict might decrease negative orientations toward Blacks. Both implicit and explicit prejudice and stereotypes were assessed. One possibility was that only explicit orientations would be influenced, given the ostensibly stubborn nature of implicit biases. However, if implicit biases are malleable, then it seemed likely that people who elect to enroll in diversity education might show decreased automatic, as well as controlled, prejudice and stereotyping.

If diversity education was successful, a second question concerned potential correlates of prejudice reduction. A key factor in multicultural training concerns awareness of one's own biases. Students are often challenged to find (and question) the ways in which they unwittingly oppress others. The key hypothesis was that learning about one's own biases might result in a decrease in implicit prejudice because of strong motives and deliberative effort expended toward becoming egalitarian (Devine, 1989). Because previous research has shown that such motives and effort can reduce explicit prejudice (Devine & Monteith, 1999), it seemed possible that similar processes might also lead to changes at the implicit level.

Experiment 1

Experimental students were enrolled in a prejudice and conflict seminar, taught by an African American male professor. Control students were enrolled in a research methods course, taught by a White female professor. Thus, this was a quasi-experiment, lacking random assignment and a fully matched control group. Assessment took place at the beginning and end of a 14-week semester. Compared with control students, experimental students were expected to show reduced prejudice and stereotyping across the two assessment periods.

The Implicit Association Test (IAT; Greenwald et al., 1998) was used to assess implicit prejudice and anti-Black stereotypes. Although introduced as an implicit attitude measure, the IAT has been extended to measure stereotypes (e.g., Rudman & Glick, in press; Rudman, Greenwald, & McGhee, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999; Rudman & Kilianski, 2000). Self-report ratings of stereotypes and the Modern Racism Scale (MRS; McConahay, 1986) were included to assess explicit intergroup orientations.

In addition, we explored potential correlates of prejudice reduction. As a result of learning about intergroup conflict, engaging in (sometimes heated) discussions, and keeping a journal documenting instances of bias (including their own), seminar students were expected to increase their awareness of prejudice, and also their motivation to counteract biases in themselves. This factor, suggestive of cognitive processes, might moderate reduction in both explicit and implicit biases (Devine, 1989; Devine & Monteith, 1999).

However, affective experiences have also been implicated in both implicit and explicit orientations. For example, liking for the professor might decrease prejudice by increasing perceptions of familiarity with Blacks (Zajonc, 1968, 1980). In addition, seminar participation might facilitate friendships with Blacks, thereby reducing prejudice through prosocial contact (Allport, 1954; Ashmore, 1969; Pettigrew, 1998). These factors, suggestive of affective processes, were included to assess their potential for reducing anti-Black biases. A priori, there was no reason to suspect that these variables would differentially covary with prejudice reduction at the implicit or explicit level.

Method

Subjects

Forty-seven volunteers (17 men and 30 women) participated in exchange for course credit. Of these, 30 were experimental students, and 17 were control students. The experimental and control groups were similar in age (both Ms = 22 years). The experimental group consisted of 18 Whites (60%), 9 Blacks (30%), and 3 others (10%). The control group consisted of 10 Whites (59%), 3 Blacks (17%), and 4 others (24%). All students completed both experimental sessions.

Stimulus Materials and Procedure

Implicit measures. The appendix shows the stimuli used in the prejudice and stereotype IATs. Each IAT used 7 White male names (e.g., *JOHN*) and 7 Black male names (e.g., *RASHAN*) as the target concepts (columns 1–2). The *prejudice IAT* used the pleasant and unpleasant meaning words shown in columns 3–4. The *stereotype IAT* used the negative attributes associated with Blacks (e.g., *lazy, hostile*) and positive attributes associated with Whites (e.g., *ambitious, calm*) that are shown in columns 5–6, respectively. Thus, the stereotype IAT captured *evaluative* stereotypes—beliefs that reflect negatively on one group, but positively on the other (Rudman et al., 2001). As a result, the prejudice and stereotype measures were conceptually similar (Rudman et al., 1999).

The IAT is administered in seven blocks, described here with materials used for the stereotype IAT. White versus Black male names are the target concepts, and positive versus negative is the attribute dimension. Subjects respond to target concepts and attributes by pressing designated right and left keys on a computer keyboard. Subjects' tasks are as follows. Block 1: They distinguish White names from Black names. For example, they might press the left key for White names and the right key for Black names. Block 2: They distinguish positive versus negative attributes. For example, they might press the left key for ambitious and the right key for lazy. Block 3: For practice, they respond to White names and positive attributes with the left key and Black names and negative attributes with the right key (abbreviated as Black+lazy). Block 4: They repeat Block 3 as a critical block. Block 5: They again distinguish positive versus negative attributes, with responses reversed. Block 6: For practice, they respond to Black names and positive attributes with the right key and White names and negative attributes with the left key (abbreviated as White+lazy). Block 7: They repeat Block 6 as a critical block. The IAT effect is computed by subtracting the mean response latency for performing the stereotype compatible task (Block 4) from the stereotype noncompatible task (Block 7). Thus, positive difference scores reflect greater tendency to associate Blacks with negative and Whites with positive traits (i.e., implicit stereotypes). The order in which subjects perform the critical blocks is counterbalanced across subjects.[2] This procedure is identical for the prejudice IAT, with pleasant versus unpleasant words replacing the positive and negative stereotypic attributes. In this case, the IAT effect represents greater tendency to associate Blacks with unpleasant and Whites with pleasant words (i.e., implicit prejudice).

The IAT was administered on desktop computers.[3] Responses were assigned to the left and right forefingers (using the *A* key and *5* key on the right-side numeric keypad, respectively). IAT stimuli appeared within a white window, vertically and horizontally centered against a light gray screen background. Subjects viewed this display from a distance of approximately 65 cm. To facilitate discrimination of target concepts (Black and White male names) from the attribute stimuli, the former were presented in uppercase black letters and the latter in lowercase blue letters. The program randomly presented stimuli from the set of possible words. There were 20 trials for the practice blocks and 40 trials for each critical block.

Explicit measures. Subjects completed the Modern Racism Scale (MRS; McConahay, 1986) as an explicit prejudice measure. The MRS consists of seven items, including "Blacks are getting too demanding in their push for equal rights," scored on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). MRS scores were averaged at Time 1 and Time 2 (mean α = .86); high scores reflect more anti-Black attitudes. As a measure of stereotypes, subjects estimated the percentage of African American and White American men who possessed each of the 12 traits used in the implicit measure (Kawakami et al., 1998). The difference between trait endorsement for Blacks and Whites for each trait was computed such that high scores reflected more explicit stereotyping. These difference scores were averaged at Time 1 and Time 2 to form the negative Black (mean α = .88) and positive White (mean α = .91) indexes. These indexes were related at Time 1 and Time 2, respectively, rs(45) = .81 and .73, ps < .001. Therefore, they were averaged to form the *stereotype* index. This index served as an explicit counterpart for the stereotype IAT.

Procedure. One to four volunteers participated individually in separate cubicles. To ensure anonymity while allowing us to match their responses across the two assessment phases, students generated their own identification number (based on a combination of digits from their social security and telephone numbers). For each assessment phase, students completed the two IATs (in counterbalanced order), as well as the explicit measures of prejudice and stereotypes. The procedural variables (counterbalancing of IATs and block order within each IAT) did not have significant effects on the findings at either Time 1 or Time 2. The average time between assessment phases was 9 weeks.

In addition, experimental students completed a measure at Time 2 designed to explore correlates of prejudice and stereotype change. The measure began with the statement, "Participating in Psychology 375 (Prejudice and Conflict) has . . ." Eight items then followed, to which students responded on 5-point scales with responses ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Three items pertained to students' increased awareness of and motivation to counteract prejudice ("Made me realize African Americans still face a lot of prejudice and discrimination"; "Opened my eyes to my own potential for biases and prejudice"; and "Made me want to work harder at overcoming my own prejudices"). These items were combined to form the *cognitive* index (α = .95). Three items assessed students' evaluation of the professor and the course ("Exposed me to an influential professor"; "Primarily been a positive, enriching experience"; and "Primarily been a negative [e.g., annoying, frustrating, or boring] experience" [reverse coded]). These items were combined to form the *evaluative* index (α = .93). Finally, two items assessed the extent to which the seminar facilitated positive contact with out-group members ("Allowed me to make friends with people outside my ethnic group"; "Allowed me to get to know more people outside my ethnic group"). These items were combined to form the *contact* index, $r(28)$ = .68, p < .001.

[2] Nonorthogonally, key assignment in Block 2 is also counterbalanced. For example, subjects who perform the White+lazy task first also press the left key for *lazy* and the right key for *ambitious*.

[3] The IAT program was written by Shelly Farnham at the University of Washington.

Results

Initial Analyses

We followed standard procedures for analyzing IAT data (Greenwald et al., 1998).[4] Because of the quasi-experimental nature of the investigation, it was important to establish that experimental and control students did not differ in their Time 1 assessments. Analyses confirmed no reliable between-group differences at Time 1, all $ts(45) < 1.61$, *ns*. Thus, students enrolled in the prejudice and conflict seminar were not less prejudiced than were control students at the beginning of the investigation.

Relationships Among Implicit and Explicit Measures

Table 1 shows the relations among the IAT and explicit measures for the combined sample. Time 1 correlations are shown above the diagonal. Time 2 correlations are shown below the diagonal. For both assessment phases, the pattern of correlations was similar. First, the prejudice and stereotype IATs were positively related, showing that implicit prejudice covaried with evaluative racial stereotypes (see also Rudman et al., 1999). Second, prejudice IAT and MRS scores showed positive covariation. That is, subjects who showed automatic anti-Black evaluation also reported prejudiced beliefs. Third, MRS and explicit stereotype scores were positively related. Finally, Table 1 shows the temporal stability coefficients for each measure on the diagonal (in italics). As can be seen, the IATs showed temporal stability reliability that was lower, but comparable, to that of the explicit measures, in support of the method's psychometric soundness.

Intergroup Orientation Change

In Experiment 1 we examined whether changes in implicit and explicit prejudice and stereotypes would occur across time for prejudice and conflict seminar students. By contrast, control students should not show orientation changes. Because the seminar's effect might be expected to differ for Black versus non-Black students, we eliminated African Americans from these (and all subsequent) analyses. The remaining sample consisted of 21 experimental and 14 control students.

Table 1
Relationships Among Implicit and Explicit Measures (Experiment 1)

Measure	Prejudice IAT	Stereotype IAT	MRS	Stereotype index
Prejudice IAT	*.50***	.41**	**.36***	**.19**
Stereotype IAT	.30*	*.48***	**.18**	**.14**
MRS	**.39****	**.09**	*.60***	.40**
Stereotype index	**.22**	**.04**	.31*	*.62***

Note. Correlations ($N = 47$) were computed by using raw latency difference scores. Correlations with transformed latencies were similar. The top matrix shows the correlations for Time 1; the bottom matrix shows the correlations at Time 2. Correlations between implicit and explicit measures are printed in bold. On the diagonal, in italics, are the temporal stability coefficients for each measure. IAT = Implicit Association Test; MRS = Modern Racism Scale.
$^{*}p < .05$. $^{**}p < .01$.

Table 2 shows the mean change scores and their effect sizes, separately for experimental and control students. These were computed such that high scores reflected "positive changes" for all measures (i.e., a decrease in prejudice and stereotyping). As can be seen, experimental students showed positive change scores, whereas control students showed negative change scores, with the exception of the explicit stereotype index. We first conducted t tests to determine whether changes in each measure were significantly different from zero. Experimental students showed decreased implicit prejudice and stereotyping over time, both $ts(20) > 3.60$, $ps < .01$. They also showed decreased explicit prejudice (MRS) and stereotyping, both $ts(20) > 3.25$, $ps < .01$. Consistent with expectations, the control group did not show significant change on any dependent measure, all $ts(13) < 1.45$, $ps > .16$.

The last column in Table 2 shows the between-groups effect sizes, estimating the effect of the seminar on intergroup orientation change. These effect sizes were large and similar in magnitude at the implicit and explicit levels. Indeed, between-groups differences were reliable for each dependent measure, all $ts(33) > 2.21$, $ps < .05$.

Factors Associated With Orientation Change

Non-Black students enrolled in the prejudice and conflict seminar showed a significant decrease in their implicit and explicit prejudice and stereotype scores across administrations. In the next set of analyses we explored potential cognitive and affective correlates of orientation changes for these students. All measures were scored so that positive relations were expected among them.

Table 3 shows the results. As can be seen, the cognitive index was positively associated with changes in explicit prejudice and stereotyping. Students who reported that the seminar increased their awareness of and motives to overcome their own biases also showed reduced MRS and stereotyping scores over time. Surprisingly, this index was only weakly related to changes at the implicit level. Instead, Table 3 shows that the evaluative index covaried positively with these changes. Students who evaluated the professor and the seminar favorably also showed reduced implicit prejudice and stereotyping scores over time. However, this index was weakly related to changes at the explicit level. Finally, the contact index showed weak associations with all change score measures, with the exception of the stereotype IAT. Students who reported making friends with out-group members also tended to show decreased implicit stereotyping, $r(19) = .41$, $p < .07$. The generally null finding for this index may be due to a lack of variability, as relatively few students reported prosocial contact with out-group members as a result of the seminar. The mean for the contact index (3.33) was significantly lower than the means for the evaluative (4.27) and cognitive (4.21) indexes, both $ts(20) > 3.40$, $ps < .01$.

[4] The first two trials of every block were eliminated because of their typically long latencies. Latencies less than 300 ms or greater than 3,000 ms were recoded as 300 and 3,000, respectively. Error trials were included in all analyses ($M = 6\%$). Latencies were initially log-transformed to normalize the distribution. However, the results were sufficiently similar to those using raw latencies in that all analyses reported are based on the untransformed latencies.

Table 2
Changes in Orientations for Experimental and Control Students (Experiment 1)

Measure	Experimental group ($n = 21$)	Control group ($n = 14$)	Pooled *SD*	Between-groups effect size
Implicit				
Prejudice	153 (**d = .74**)	−51 (**d = −.24**)	207.81	**0.98**
Stereotype	118 (**d = .86**)	−48 (**d = −.35**)	136.60	**1.22**
Explicit				
MRS	0.28 (**d = .47**)	−0.27 (**d = −.46**)	0.59	**0.93**
Stereotype	2.55 (**d = .91**)	0.45 (**d = .16**)	2.81	**0.75**

Note. Only non-Black students were used in these analyses ($N = 35$). All measures are difference scores, computed so that positive scores reflect a decrease in prejudice and stereotyping from Time 1 to Time 2. Implicit Association Test measures are based on a millisecond index. Effect sizes (printed in bold) are Cohen's *d*. Within-group effect sizes were computed by dividing experimental and control subjects' difference score means by the pooled standard deviation. Between-groups effect sizes were computed by subtracting control subjects' effect size from experimental subjects' effect size. Conventional small, medium, and large effect sizes for *d* are .2, .5, and .8, respectively (Cohen, 1988). MRS = Modern Racism Scale.

In summary, Experiment 1's results unexpectedly showed that explicit and implicit orientation changes were better related to factors suggestive of cognitive and affective processes, respectively. However, the indexes themselves were not independent. The relationship between the cognitive and evaluative indexes was significantly positive, $r(19) = .44$, $p < .05$, and each was positively (albeit weakly) related to the contact index, both $rs(19) < .24$, *ns*. Thus, the constructs assessed by the affective and cognitive indexes were related, yet were differentiable in the way in which they covaried with changes in implicit versus explicit prejudice and stereotyping.

Finally, a check on the relations among *change score* indexes showed that the prejudice and stereotype IAT change scores covaried, $r(19) = .41$, $p < .07$, as did the explicit MRS and stereotype change scores, $r(19) = .55$, $p < .01$. In support of their discriminant validity, relations between the implicit and explicit change scores were weakly positive, all $rs(19) < .15$, *ns*.

Discussion

Experiment 1's focal finding was that prejudice and conflict seminar students showed less anti-Black biases at the end of the semester, compared with the beginning. Moreover, they did so at both the implicit and explicit levels. That is, students exposed to coursework and class discussions designed to foster respect for diversity showed a significant reduction in both their prejudice and stereotype IAT scores. In addition, these students showed reduced self-reported prejudice and stereotyping. By contrast, control students did not show significant change in either implicit or explicit orientations. Although this demonstration was quasi-experimental in nature, it provides promising evidence that participating in a seminar concerned with race-related issues, led by an African American professor, may have generalized positive effects on both implicit and explicit prejudice and stereotypes. The primary objective of Experiment 2 was to replicate this central finding because it suggests the positive benefits that diversity training may have on volunteers. Moreover, this finding contradicts conceptualizing implicit biases as intractable.

Intriguingly, Experiment 1's results suggested that prejudice and conflict seminar students' changes at the automatic and controlled level were distinguishable in at least two respects. First, they were only weakly (albeit positively) correlated. Thus, reduced implicit bias did not necessarily follow from reduced explicit bias. Second, implicit and explicit changes reliably covaried with different factors. Explicit change scores were associated with a cognitive variable (increased awareness of and motives to counteract own biases). This finding extends prior work in prejudice reduction (Devine & Monteith, 1999) to a real-world situation, whereby students gain insight into their own biases via multicultural training and work to overcome these. By contrast, implicit changes covaried with affective variables, including favorable attitudes toward the professor and, in the case of implicit stereotyping, prosocial contact with out-group members. These findings were somewhat surprising because the cognitive and affective variables might be expected to relate to both explicit and implicit modifications.

In Experiment 2, we sought to replicate and extend this pattern by better determining how the professor and the course reduced implicit anti-Black attitudes and beliefs. To that end, we added a third affective variable—seminar students' reports of feeling less

Table 3
Cognitive and Affective Correlates of Experimental Students' Orientation Change Scores (Experiment 1)

Change score measure	Cognitive index[a]	Evaluative index[b]	Contact index[c]
Prejudice IAT	.04	.48*	.04
Stereotype IAT	.13	.46*	.41
MRS	.48*	.11	.07
Stereotype index	.54*	.06	.10

Note. Only non-Black students were used in these analyses ($N = 21$). Change score measures were computed so that high scores correspond to greater reduction in prejudice and stereotypes. IAT = Implicit Association Test; MRS = Modern Racism Scale.
[a] Increased awareness of discrimination against African Americans and motives to overcome prejudice in oneself. [b] Positive evaluation of the professor and the prejudice and conflict seminar. [c] Increased friendships and acquaintances with out-group members.
* $p < .05$.

threatened by out-group members as a result of participating in the course. This index was suggested by the correspondence shown between automatic prejudice and activation of a neural substrate associated with emotional conditioning, including fear-based responses (Amodio et al., 2000; Phelps et al., 2000). Because the seminar provides ongoing interactions with African Americans, in the form of the professor and fellow students, non-Black students may develop increased feelings of comfort with this group. That is, the seminar may provide a context for "unlearning" anxiety associated with African Americans. If so, reductions in threat perceptions and implicit prejudice should positively correlate, which coincides with past neurological findings.

Finally, we improved Experiment 2's design in two ways. First, whereas Experiment 1 used only one implicit method, Experiment 2 used semantic priming to measure stereotypes (the lexical decision task [LDT]; Wittenbrink et al., 1997) and associative categorization to measure attitudes (Experiment 1's prejudice IAT). This change allowed us to test the generalizability of Experiment 1's results. Second, we added a control group consisting of students enrolled in a lecture course taught by the same African American professor. This change allowed us to determine whether the professor was sufficient for reducing anti-Black biases, or whether the prejudice and conflict seminar was also required. It also provided a larger sample of African Americans with which to examine the implicit methods' known groups validity (see also Fazio et al., 1995; Greenwald et al., 1998; Rudman et al., 1999).

Experiment 2

As in Experiment 1, experimental students were enrolled in a prejudice and conflict seminar, taught by the same African American male professor. Control students were enrolled either in a large lecture course taught by the identical professor or in a research methods course taught by a White female professor. Assessment took place at the beginning and end of a 14-week semester. Only experimental students (not control students) were expected to show diminished prejudice across the two assessment periods.

Method

Subjects

One hundred and nineteen volunteers (33 men and 86 women) participated in both experimental sessions in exchange for course credit.[5] Of these, 28 were experimental students and 91 were control students (62 from the lecture course and 29 from the research methods course). The experimental and control groups were similar in age (both $Ms = 22$ years). The experimental group consisted of 5 Blacks (18%) and 23 non-Blacks (82%). The control group consisted of 28 Blacks (31%) and 63 non-Blacks (69%).

Stimulus Materials and Procedure

Implicit measures. The appendix shows the stimulus words used in the prejudice IAT and stereotype LDT. Each measure used 7 White male names (e.g., *JOHN*) and 7 Black male names (e.g., *RASHAN*) as group tokens (columns 1–2). These served as either the target concepts (IAT) or primes (LDT). The IAT used the pleasant and unpleasant meaning words shown in columns 3–4. The LDT used the negative attributes associated with Blacks (e.g., *lazy, hostile*) and positive attributes associated with Whites (e.g., *ambitious, calm*) that are shown in columns 5–6, respectively.

The IAT was administered and scored exactly as in Experiment 1, with positive difference scores reflecting greater tendency to associate Blacks with unpleasant and Whites with pleasant evaluations (i.e., implicit prejudice). The LDT was administered and scored following prior research (Wittenbrink et al., 1997). Subjects' task was ostensibly to differentiate words from nonwords (e.g., letter strings). However, the true purpose was to examine whether particular words are recognized faster than other words when preceded by a prime. After receiving computerized instructions and 10 practice trials, subjects performed the critical trials. For each trial, a warning signal (+) appeared in the center of the CRT screen for 500 ms. This was followed by a 15-ms exposure to a prime, a visual mask (***) for 200 ms, and then the target word (shown in the appendix). Subjects indicated whether the target stimulus was a word or a nonword by pressing the *Q* or *P* keys on a keyboard, respectively. Primes were of three types: a neutral prime (*XYZX*), the word *BLACKS*, or the word *WHITES*. Trials in which the prime was neutral formed the baseline latencies for all stimuli. Facilitation scores—the difference between baseline and critical trials—were then computed for each trial type (e.g., *BLACKS*/Negative, *WHITES*/Positive, *BLACKS*/Positive, and *WHITES*/Negative) such that high scores indicate faster recognition compared with baseline. The critical dependent measure was a facilitation contrast score that represented, in a single index, implicit evaluative stereotyping. Specifically, high scores indicate greater facilitation for recognizing negative Black words (e.g., *lazy*) when primed with *BLACKS* versus *WHITES*, and greater facilitation for recognizing positive White words (e.g., *ambitious*) when primed with *WHITES* versus *BLACKS* (Wittenbrink et al., 1997).[6]

Explicit measures. Subjects completed thermometer measures that separately assessed attitudes toward Blacks and Whites. Each was scored on a scale ranging from 0 (*extremely cold*) to 100 (*extremely warm*). The difference between these was computed such that high scores indicated more negative attitudes toward Blacks than Whites (i.e., explicit prejudice). Because the *thermometer* index is a difference score, it may serve as a better explicit counterpart to the IAT (compared with Experiment 1's MRS). Subjects also completed Experiment 1's stereotype measure. As in Experiment 1, the negative Black and positive White difference score indexes were related at Time 1 and Time 2, respectively, $rs(117) = .58$ and .43, $ps < .001$. They were subsequently averaged to form the stereotype index, which served as an explicit counterpart to the LDT.

Procedure. Experiment 2 followed Experiment 1's protocol. For each assessment phase, subjects completed the implicit measures (in counterbalanced order), as well as the explicit measures of prejudice and stereotypes. The procedural variables (counterbalancing of implicit measures and block order within the IAT) did not have significant effects on the findings at either Time 1 or Time 2. The average time between assessment phases was 9 weeks.

At Time 2, experimental students also completed Experiment 1's cognitive ($\alpha = .86$) and evaluative ($\alpha = .85$) indexes and the two-item contact measure, $r(26) = .62, p < .001$. In addition, they completed two items unique to Experiment 2: "[The seminar] allowed me to feel less threatened by people

[5] Thirteen students from the lecture course did not return for the second session because they had dropped the class. Their data are not included in this article.

[6] Following Wittenbrink et al. (1997), positive words stereotypic of Blacks (e.g., *athletic* and *musical*) and negative words stereotypic of Whites (e.g., *boring* and *stiff*) were also used. However, we found, as Wittenbrink et al. did, that subjects did not possess a positive stereotype of Blacks or a negative stereotype of Whites at the implicit level. We therefore treated the positive Black and negative White stereotypic attributes as filler items in the present research. The contrast score that we used as the single stereotyping index corresponds to Wittenbrink et al.'s *alternative stereotyping* index, which was found to correlate with explicit measures of prejudice.

outside my ethnic group" and "[The seminar] made me feel more comfortable with people outside my ethnic group." Responses on the measure ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). These items were averaged to form the *fear reduction* index, $r(26) = .82$, $p < .001$.

Results and Discussion

Initial Analyses

We followed standard procedures for analyzing IAT (Greenwald et al., 1998) and LDT data.[7] In Experiment 2 we used two control groups: one taught by the African American professor, the other by a White female professor. We first determined that there were no reliable group differences at Time 1 on the implicit and explicit measures of prejudice and stereotyping for control students, all $ts(89) < 1.12$, *ns*. At Time 2, if the African American professor was sufficient for reducing bias, then we should expect the first group to score lower than the second group on these measures. However, analyses showed no significant differences, all $ts(89) < 1.75$, *ns*. In summary, the control groups scored similarly on all measures at both assessment phases. They were therefore combined to form a single control group.

As in Experiment 1, it was important to establish that students enrolled in the prejudice and conflict seminar were not less prejudiced than were control students at the start of the investigation. Analyses of their Time 1 measures supported this assumption, all $ts(117) < 1.37$, *ns*.

Table 4
Summary Statistics for Implicit and Explicit Measures (Experiment 2)

Measure	Non-Blacks (n = 86)	Blacks (n = 33)	Pooled *SD*	Effect size
Time 1				
Implicit				
Prejudice IAT	149.25	−25.13	181.29	**0.95**
Stereotype LDT	10.47	−4.89	86.00	**0.17**
Explicit				
Thermometer index	3.31	−15.00	16.33	**1.12**
Stereotype index	6.33	3.02	9.61	**0.34**
Time 2				
Implicit				
Prejudice IAT	182.99	51.16	181.73	**0.72**
Stereotype LDT	3.05	−1.45	50.00	**0.03**
Explicit				
Thermometer index	3.55	−17.12	20.30	**1.01**
Stereotype index	5.15	0.89	11.47	**0.37**

Note. Implicit measures are based on a millisecond index. The Implicit Association Test (IAT) and thermometer measures are difference scores. The lexical decision task (LDT) and stereotype indexes are averaged difference (i.e., contrast) scores. In each case, high scores reflect greater prejudice or stereotyping. Effect sizes for the difference between non-Blacks and Blacks (printed in bold) are Cohen's *d*. These between-groups effect sizes were computed by subtracting control subjects' mean from experimental subjects' mean, and dividing by the pooled standard deviation. Conventional small, medium, and large effect sizes for *d* are .2, .5, and .8, respectively (Cohen, 1988).

Black and Non-Black Group Differences

Table 4 displays summary statistics for Experiment 2's measures, separately for Black and non-Black students, at Time 1 and Time 2. For each measure, high scores reflect greater anti-Black prejudice and stereotyping. The primary goal was to examine the implicit measures' known groups validity. Thus, non-Blacks were expected to show higher scores than Blacks. Comparison tests supported this hypothesis for the prejudice IAT at each assessment phase, both $ts(117) > 3.71$, $ps < .001$, but not for the stereotype LDT, both $ts(117) < 1.00$. In addition, non-Blacks scored reliably higher than Blacks on the thermometer measure at each phase, both $ts(117) > 5.55$, $ps < .001$. However, there were no group differences on the stereotype index at either Time 1 or Time 2, both $ts(117) < 1.81$, $ps > .08$. In summary, there were reliable between-groups differences for both prejudice measures, but not for either stereotyping measure. The IAT's known groups validity is consistent with past findings (e.g., Greenwald et al., 1998; Rudman et al., 1999). Although the LDT did not distinguish between groups, the explicit stereotype index also suggested similar possession of stereotypes for Blacks and non-Blacks.

Relationships Among Implicit and Explicit Measures

Table 5 shows the relations among the implicit and explicit measures for the combined sample. The top matrix shows the relationships at Time 1; the bottom matrix shows the relationships at Time 2. We were particularly interested in whether the implicit measures would show convergence. The prejudice IAT and stereotype LDT were positively related, and reliably so at Time 1. This finding is somewhat impressive, given the two measures' methodological differences. That is, the IAT used associative categorization to index affective associations, whereas the LDT used subliminal priming to capture evaluative semantic associations (see also Brauer et al., 2000; Rudman & Kilianski, 2000). By contrast, the explicit thermometer and stereotyping indexes were positively associated at both assessment phases.

With respect to implicit–explicit convergence, Table 5 shows that the prejudice IAT was related to the thermometer and stereotyping indexes at both Time 1 and Time 2. Students who showed automatic anti-Black evaluation also reported a preference for Whites and anti-Black beliefs. By contrast, the stereotype LDT was generally weakly (albeit positively) related to the explicit measures. Finally, Table 5 shows the temporal stability coefficients for each measure on the diagonal (in italics). These were significantly positive for all measures, with the exception of the stereotype LDT. When considering all measures, Table 5 reveals more evidence for the psychometric soundness of the prejudice IAT as compared with the stereotype LDT.

Intergroup Orientation Change

Following Experiment 1, change scores were computed such that positive scores reflected a reduction in bias at Time 2, compared with Time 1, for all measures. We again eliminated African Americans from these (and subsequent) analyses, resulting in a sample of 23 experimental and 63 control students.

[7] Analyses of subjects' accuracy revealed low error rates for both the IAT and the LDT (an average of 5% and 1%, respectively); error trials were included in all analyses. For each measure, latencies were log-transformed to normalize the distribution. As in Experiment 1, the results were sufficiently similar to those using raw latencies that all analyses reported are based on the untransformed latencies.

Table 5
Relationships Among Implicit and Explicit Measures (Experiment 2)

Measure	IAT	LDT	THERM	STP
Prejudice IAT	*.47***	.18*	**.42****	**.33****
Stereotype LDT	.12	*.08*	**.13**	**.15**
THERM	**.25***	.16	*.80***	.40**
STP	**.20***	.11	.43**	*.76***

Note. Correlations (N = 119) were computed by using raw latency difference scores. Correlations with transformed latencies were similar. The top matrix shows the correlations for Time 1; the bottom matrix shows the correlations at Time 2. Correlations between implicit and explicit measures are printed in bold. On the diagonal, in italics, are shown the temporal stability coefficients for each measure. IAT = Implicit Association Test; LDT = lexical decision task; THERM = thermometer index; STP = stereotype index.
* $p < .05$. ** $p < .01$.

Table 6 shows the mean change scores and their effect sizes, separately for each group. As in Experiment 1, only experimental students showed consistently positive change scores. These scores were reliably different from zero for both implicit measures and the explicit stereotyping index, all ts(22) > 2.00, ps < .05. Only the thermometer measure did not show reliable change for experimental students, t(22) < 1.00. As expected, the control group did not show significant changes over time, all ts(62) < 1.89, *ns*.

The last column in Table 6 shows the between-groups effect sizes, estimating the effect of the seminar on intergroup orientation change. These were moderate to large in magnitude, with the exception of the thermometer index. Comparison tests showed reliable group differences for the IAT, LDT, and explicit stereotype measures, all ts(84) > 2.80, ps < .01. Only the thermometer index did not distinguish between groups, t(84) < 1.00. Nonetheless, three out of four measures conformed to hypotheses, in support of Experiment 1's central finding.

In summary, Experiment 2 replicated Experiment 1 by showing a greater reduction in prejudice and stereotyping for prejudice and conflict students as compared with control students. In each experiment, this decrease was reliable for the implicit measures, irrespective of the method used (IAT or LDT). Furthermore, each experiment showed decreased explicit biases for prejudice and conflict students, with the single exception of Experiment 2's thermometer measure. These findings strongly support the hypothesis that people can "unlearn" both explicit and implicit prejudice in real-world contexts.

Factors Associated With Orientation Changes

In Experiment 2 we continued to examine potential cognitive and affective correlates of orientation changes. In Experiment 1, the cognitive index covaried with changes in explicit biases, whereas the *affective* indexes (evaluative and contact) covaried with changes in implicit biases. In Experiment 2 we also sought to replicate and extend this pattern by adding the fear reduction index. All measures were scored so that positive relations were expected among them.

Table 7 shows the results. As can be seen, they mirror Experiment 1's findings in several ways. First, the cognitive index was significantly and positively related to changes in explicit stereotyping. Students who reported greater awareness of discrimination and motives to "work hard" to counter their biases also showed decreased stereotyping at Time 2 compared with Time 1. As in Experiment 1, this index was not reliably related to changes at the implicit level. Second, the evaluative and contact indexes were positively associated with changes in implicit prejudice and stereotyping. Students who evaluated the professor and the course favorably, or who reported making friends with out-group members during the seminar, also showed less automatic prejudice and stereotypic beliefs over time. As in Experiment 1, these affective indexes were not reliably related to changes at the explicit level. New to Experiment 2, the fear reduction index also covaried significantly and positively with changes in implicit prejudice. It also showed a marginally positive relationship with changes in implicit stereotyping, $r(21)$ = .36, $p < .09$. Thus, students who reported feeling less threatened by out-group members as a result of seminar participation also showed reduced implicit prejudice and stereotyping. Consistent with the pattern shown for the evaluative and contact indexes, the fear reduction measure was not reliably related to changes in explicit orientations. In concert, these

Table 6
Changes in Orientations for Experimental and Control Students (Experiment 2)

Change score measure	Experimental group (n = 23)	Control group (n = 63)	Pooled *SD*	Between-groups effect size
Implicit				
Prejudice IAT	96 (***d*** = **.54**)	−58 (***d*** = **−.33**)	177.02	**.85**
Stereotype LDT	48 (***d*** = **.42**)	−24 (***d*** = **−.21**)	112.07	**.65**
Explicit				
Thermometer index	1.97 (***d*** = **.17**)	0.27 (***d*** = **.02**)	11.50	**.15**
Stereotype index	4.72 (***d*** = **.59**)	0.07 (***d*** = **0**)	8.00	**.59**

Note. Only non-Black students were used in these analyses (N = 86). All measures are difference scores, computed so that positive scores reflect a decrease in prejudice and stereotyping from Time 1 to Time 2. Implicit measures are based on a millisecond index. Effect sizes (printed in bold) are Cohen's *d*. Within-group effect sizes were computed by dividing experimental and control subjects' difference score means by the pooled standard deviation. Between-groups effect sizes were computed by subtracting control subjects' mean from experimental subjects' mean, and dividing by the pooled standard deviation. Conventional small, medium, and large effect sizes for *d* are .2, .5, and .8, respectively (Cohen, 1988). IAT = Implicit Association Test; LDT = lexical decision task.

Table 7
Cognitive and Affective Correlates of Experimental Students' Orientation Change Scores (Experiment 2)

Change score measure	Cognitive index[a]	Evaluative index[b]	Contact index[c]	Fear reduction index[d]
Prejudice IAT	.26	.47*	.55*	.50*
Stereotype LDT	.09	.42*	.49*	.36
Thermometer index	.25	.26	.22	.16
Stereotype index	.47*	−.12	.15	.15

Note. Only non-Black students were used in these analyses ($N = 23$). Change score measures were computed so that high scores correspond to greater reduction in prejudice and stereotypes. IAT = Implicit Association Test; LDT = lexical decision task.
[a] Increased awareness of discrimination against African Americans and motives to overcome prejudice in oneself. [b] Positive evaluation of the professor and the prejudice and conflict seminar. [c] Increased friendships and acquaintances with out-group members. [d] Reduced fear in the presence of out-group members as a result of seminar participation.
* $p < .05$.

findings support Experiment 1's intimation that implicit and explicit prejudice reduction is associated with factors suggestive of affective versus cognitive processes, respectively.

The one anomalous finding concerns the thermometer measure, which showed weak relations with all four indexes. However, the modal response for all students at Time 1 and Time 2 on this index was zero (i.e., no preference for either Blacks or Whites), resulting in a modal mean change score of zero (52% of the non-Black experimental group showed this score; see also Greenwald et al., 1998, Experiment 3). Therefore, the measure's lack of variability may have contributed to relatively low statistical power with which to find relationships.

A check on the relationships among experimental students' Time 2 indexes showed reliable covariation among the affective indexes. Students who evaluated the professor and the course favorably also reported making friends with out-group members and reduced feelings of out-group threat, *r*s(21) = .66 and .58, respectively, *p*s < .01. Furthermore, prosocial contact with out-group members was strongly related to fear reduction, $r(21) = .79$, $p < .001$. In addition, the cognitive index was significantly related to the fear reduction index, $r(21) = .42$, $p < .05$. Finally, the cognitive index was positively but weakly related to the evaluative and contact indexes, *r*s(21) = .33 and .18, respectively, *ns*.

A check on the relations among change score indexes showed a pattern consistent with Experiment 1's findings. First, the implicit prejudice and stereotype change scores positively covaried, $r(21) = .55$, $p <.01$, as did the explicit prejudice and stereotype change scores, $r(21) = .57$, $p < .01$. Second, the implicit and explicit prejudice and stereotype change scores were weakly (but positively) related, all *r*s(21) < .37, *ns*. These findings suggest that changes in intergroup biases at the implicit and explicit levels represent related but distinct events.

Comparison of LDT and IAT Measures

Experiment 2 allowed us to compare associative categorization and semantic priming techniques for assessing anti-Black orientations. In general, the prejudice IAT showed more validity than did the stereotyping LDT, including known groups validity, temporal stability, and convergence with explicit measures of intergroup orientations. The LDT results were, frankly, surprising. Although its known groups validity and temporal stability have not been previously assessed, the LDT stereotyping index used here has covaried with explicit measures of prejudice in the past (Wittenbrink et al., 1997). However, its relationship with explicit stereotypes has not previously been tested. Its lack of covariation with the present explicit stereotyping index, and its positive association with the prejudice IAT (significant at Time 1), in tandem with the positive covariation shown between LDT and IAT change scores, suggest that the stereotype LDT reflects implicit evaluation as well as beliefs—a possibility acknowledged by Wittenbrink et al.

Nonetheless, the LDT and IAT measures' performance was comparable in at least two respects. First, each measure was sensitive to reduced bias on the part of prejudice and conflict seminar students relative to control students. Second, for each measure, these changes were positively linked to affective measures (including prosocial contact and increased comfort with out-group members). Indeed, the fact that LDT changes corresponded to these affect-based measures in nearly identical fashion as compared with the prejudice IAT suggests, again, that the index used here assesses implicit prejudice as well as stereotyping.

General Discussion

In the two experiments reported here, prejudice and conflict seminar students showed decreased anti-Black biases at the end of the semester as compared with at the beginning of the semester. These findings represent the first known efforts to evaluate the effectiveness of multicultural training on student volunteers. Although the investigations were quasi-experimental, they provide promising evidence that participating in a seminar concerned with race-related issues, led by an African American professor, may have positive effects on intergroup orientations.

Importantly, these effects were evident when both self-report and automatic methods were used, despite the fact that implicit orientations are thought to be intractable. Contrary to this assumption, students who voluntarily enrolled in diversity education showed a significant reduction in their implicit prejudice and stereotype scores, compared with control students. Our use of two implicit techniques in Experiment 2 enhances confidence in the generalizability of the findings and suggests that multicultural education can modify people's attitudes and beliefs at the automatic level.

How critical was the African American professor to these findings? Although a majority of Experiment 2's control group (68%) was also taught by the same professor, these students did not show reduced implicit or explicit scores across the two assessment phases. Thus, the content of the prejudice and conflict seminar, as well as its relatively intimate atmosphere, may have fostered the openness and appreciation for diversity necessary to enable the unlearning of implicit and explicit biases.

Limitations of the Research

Although these results suggest that diversity education is beneficial, they are limited in at least three respects. First, we used the same instructor for the prejudice and conflict course in each investigation. Without investigating students enrolled in the identical course taught by a different (e.g., White) professor, the effect of the seminar's content alone cannot be known. Second, the long-term effects of the prejudice and conflict seminar were not assessed. Future research is needed to determine whether the observed decreases in implicit and explicit biases are temporary or stable. Third, the data are limited by their quasi-experimental nature. Although students who enrolled in the prejudice and conflict seminar did not show lower levels of bias than did control students at the start of each experiment, they may have differed from control students in other important ways. In particular, their internal standards to be nonprejudiced may have been stronger (Plant & Devine, 1998)—a possibility that may have initially led them to enroll in the course. In essence, students volunteered to be in the experimental group. Although this fact prevents scientific control, people in the real world generally select themselves into situations that they believe will affect their behavior in desirable ways (Snyder & Ickes, 1985). In this respect, people who volunteer for diversity education may be particularly receptive to the seminar's message (i.e., ready and willing to change their beliefs).

How Do Explicit and Implicit Orientations Change?

In two experiments, changes in experimental students' explicit and implicit orientations covaried with factors suggestive of cognitive and affective processes, respectively. Consistent with past research (Devine & Monteith, 1999), insight into one's own biases and motives to be nonprejudiced were linked to reduced explicit prejudice. As a general rule, awareness of bias is critical for countering mental contamination (e.g., Wilson & Brekke, 1994). This has been problematic in race relations because it is difficult for many non-Blacks to admit or "realize" their prejudicial attitudes and beliefs (Gaertner & Dovidio, 1986). The present findings indicate that those who did so in the context of diversity training, and who were concerned with becoming egalitarian, responded by reducing their explicit prejudice and stereotypes. However, the fact that this index was not reliably associated with implicit orientation changes suggests that "something else" may be needed to affect automatic biases.

The something else appears to be affective in nature. Prejudice and conflict seminar students who evaluated the professor and the course positively, who made friends with out-group members, and who reported feeling less threatened by out-group members also showed decreased implicit prejudice and anti-Black stereotypes. Because evaluation, making friends, and fear reduction are suggestive of affective processes, one route to implicit orientation change may be emotional. Indeed, neurological evidence suggests that implicit prejudice reflects affect-based responses toward social groups (Amodio et al., 2000; Phelps et al., 2000). As a result, emotional reconditioning may be an effective means of reducing automatic biases (cf. Kawakami et al., 2000).

Of course, the correlates of explicit and implicit change can be differentiated on other dimensions besides cognitive and affective. In particular, awareness of bias and motives to become more egalitarian may represent an intentional learning process, one that is directly gained as a result of the course. That is, the cognitive index may have captured the explicit message of the seminar. By contrast, affective reactions may represent an incidental learning process, one that is indirectly gained by participating in the course. It has been argued that automatic biases stem from indirect learning (Devine, 1989). If people unintentionally acquire implicitly prejudicial attitudes and beliefs, why would it be necessary for them to deliberately unlearn them? Indeed, people motivated to "try hard" not to be prejudiced were unable to change their IAT scores (Kim & Greenwald, 1998). By contrast, people briefly exposed to positive Black exemplars (Dasgupta & Greenwald, 2001) or primed with counterstereotypic mental imagery (Blair et al., 2001) did show reduced implicit biases through processes that seem to be relatively indirect.

The dimensions distinguishing the cognitive and affective indexes need not be mutually exclusive. Together, though, they support a matching hypothesis such that attitude change may be most efficient when the persuasion route matches the targeted attitude's characteristics (Edwards, 1990; Edwards & von Hippel, 1995; Fabrigar & Petty, 1999). The present findings, although speculative, suggest that explicit intergroup orientations may be linked more to cognitive or direct processes, whereas implicit intergroup orientations may be linked more to affective or indirect processes. Future research is necessary to determine the extent to which these processes contribute to orientation change at each level, but the present data tentatively point to distinct types of learning.

This is not to suggest that the two processes are independent. For example, cognitive factors may have led people to pursue other opportunities for change (e.g., emotional reconditioning). Indeed, in each experiment, the cognitive and affective indexes positively covaried. Thus, seminar students who gained insight into their own biases may have liked the professor more and sought friendships with Black students, which may have led to increased comfort with African Americans. Of course, these relationships could work in reverse; the greater point is that the two processes are likely to work hand in hand to promote changes in automatic biases.

Finally, our research underscores the importance of using implicit techniques to detect biases that are not likely to be disclosed to others, or even to one's self. The pervasive evidence for implicit prejudices has been disturbing to uncover, but their origins are not mysterious, nor are they intractable. The present findings suggest that, for volunteers, educational forums designed to promote appreciation for diversity, friendships with out-group members, and insight into one's own prejudice and stereotypes can enable the unlearning of both implicit and explicit intergroup biases—a possibility that should inspire cautious optimism for researchers and educators alike.

References

Allport, G. W. (1954). *The nature of prejudice.* New York: Addison-Wesley.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2000). *The emotional and physiological components of race bias: Individual differences in attention- and emotion-modulated startle eyeblink response.* Manuscript submitted for publication.

Ashmore, R. D. (1969). Intergroup contact as a prejudice-reduction technique: Experimental examination of the shared-coping approach and four alternative explanations. *Dissertation Abstracts International, 31,* 05, 2949B.

Banks, J. A. (1995). Multicultural education: Its effects on students' racial and gender role orientation. In J. A. Banks & C. M. Banks (Eds.), *Handbook of research on multicultural education* (pp. 617–627). New York: Macmillan.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford Press.

Beal, D. J., O'Neal, E. C., Ong, J., & Ruscher, J. B. (2000). The ways and means of interracial aggression: Modern racists' use of covert aggression. *Personality and Social Psychology Bulletin, 26,* 1225–1238.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues, 55,* 687–706.

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

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 11, pp. 1–52). Hillsdale, NJ: Erlbaum.

Brauer, M., Wasel, W., & Niedenthal, P. (2000). Implicit and explicit components of prejudice. *Review of General Psychology, 4,* 79–101.

Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.

Cohen, J. (1988). *Statistical power for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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.

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

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814.

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

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 Press.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. (2000). *Exploring the relationship between implicit and explicit prejudice: The role of motivations to respond without prejudice*. Manuscript submitted for publication.

Dijksterhuis, A., & van Knippenberg, A. (1996). The knife that cuts both ways: Facilitated and inhibited access to traits as a result of stereotype activation. *Journal of Experimental Social Psychology, 32,* 271–288.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. M. 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. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11,* 315–319.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology* (Vol. 4, pp. 175–197). Oxford, England: Blackwell.

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

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

Eberhardt, J. L., & Fiske, S. T. (1998). *Confronting racism: The problem and the response.* Thousand Oaks, CA: Sage..

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology, 59,* 202–216.

Edwards, K., & von Hippel, W. (1995). Hearts and minds: The priority of affective versus cognitive factors in person perception. *Personality and Social Psychology Bulletin, 21,* 996–1011.

Fabrigar, L. R., & Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Personality and Social Psychology Bulletin, 25,* 363–381.

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.

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. 61–90). Orlando, FL: Academic Press.

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

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

Hewstone, M. (1996). Contact and categorization: Social psychological interventions to change intergroup relations. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 323–368). New York: Guilford Press.

Johnston, L., & Hewstone, M. (1992). Cognitive models and stereotype change: III. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28,* 360–386.

Jones, J. M. (1972). *Prejudice and racism.* New York: McGraw-Hill.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788.

Kawakami, K., Dion, K. L., & Dovidio, J. F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin, 24,* 407–416.

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

Kim, D., & Greenwald, A. G. (1998, May). *Voluntary controllability of implicit cognition: Can implicit attitudes be faked?* Paper presented at the 70th annual meeting of the Midwest Psychological Association, Chicago.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits and behaviors: A parallel-constraint satisfaction theory. *Psychological Review, 103,* 284–308.

McCauley, C., Wright, M., & Harris, M. E. (2000). Diversity workshops on campus: A survey of current practice at U.S. colleges and universities. *College Student Journal, 34,* 100–114.

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. 91–126). Orlando, FL: Academic Press.

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

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.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999).

Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77,* 167–184.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49,* 65–85.

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

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811–832.

Plant, E. A., & Devine, P. G. (2000). *Responses to other-imposed pro-Black pressure: Acceptance or backlash?* Manuscript submitted for publication.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114,* 363–375.

Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology, 31,* 493–517.

Rudman, L. A., & Glick, P. (in press). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues.*

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

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17,* 437–465.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26,* 1315–1328.

Schofield, J. W. (1986). Causes and consequences of the colorblind perspective. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 231–254). Orlando, FL: Academic Press.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations.* Cambridge, MA: Harvard University Press.

Sears, D. (1988). Symbolic racism. In P. Katz & D. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53–84). New York: Plenum.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Hernstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., pp. 739–811). New York: Wiley.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1, pp. 883–948).

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45,* 961–977.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34–52.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 446–496). New York: Oxford University Press.

Wilder, D. A. (1984). Intergroup contact: The typical member and the exception to the rule. *Journal of Experimental Social Psychology, 20,* 177–194.

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

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

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology, 78,* 635–654.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9* (Monograph Suppl. 2, Pt. 2).

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–171.

Appendix

Stimulus Words for Implicit Prejudice and Stereotype Measures

Group tokens		Prejudice IAT		Stereotype LDT	
White name	Black name	Pleasant word	Unpleasant word	Negative Black trait	Positive White trait
JOHN	RASHAN	sunshine	filth	lazy	ambitious
BRAD	MALIK	smile	death	shiftless	industrious
PAUL	DARNEL	angel	devil	unemployed	successful
BRIAN	TYRCEL	luck	slime	hostile	calm
PETER	JAMAL	rainbow	cancer	dangerous	trustworthy
ROBERT	LEVON	paradise	hell	threaten	ethical
ANDREW	GEROME	fortune	poison	violent	lawful

Note. In Experiment 1, the stereotype IAT used the same stimuli shown for the stereotype LDT. Stimulus words were adopted from past research (Greenwald, McGhee, & Schwartz, 1998; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Wittenbrink, Judd, & Park, 1997). IAT = Implicit Association Test; LDT = lexical decision task.

Received July 18, 2000
Revision received February 26, 2001
Accepted February 26, 2001 ■