Inducing Hypocrisy to Reduce Prejudicial Responses among Aversive Racists

Article *in* Journal of Experimental Social Psychology · January 2002

Impact Factor: 2.29 · DOI: 10.1006/jesp.2001.1484

3 authors, including:

Leanne S. Son Hing
University of Guelph

12 PUBLICATIONS **483** CITATIONS

SEE PROFILE

Journal of Experimental Social Psychology **38,** 71–78 (2002)
doi:10.1006/jesp.2001.1484, available online at http://www.idealibrary.com on IDEAL®

Inducing Hypocrisy to Reduce Prejudicial Responses among Aversive Racists

Leanne S. Son Hing, Winnie Li, and Mark P. Zanna

University of Waterloo, Waterloo, Ontario, Canada

Received June 30, 2000; revised March 15, 2001; accepted March 15, 2001; published online June 11, 2001

Might a hypocrisy induction procedure reduce prejudicial behavior among aversive racists? We identified aversive racists as individuals low in explicit prejudice but high in implicit prejudice toward Asians. Results revealed that aversive racists, but not truly low prejudiced participants (i.e., those low in both explicit and implicit prejudice), responded to a hypocrisy induction procedure with increased feelings of guilt and discomfort, compared to those in a control condition. Furthermore, aversive racists, but not low prejudiced participants, responded to a hypocrisy induction procedure with a reduction in prejudicial behavior. These results suggest that consciousness raising might play an important role in motivating aversive racists to reduce their prejudicial behavior.

Key Words: prejudice; discrimination; prejudice reduction; aversive racism; implicit attitudes; hypocrisy induction; Asians.

As a new millennium begins, evidence of discrimination based on race and gender is still found in many field and laboratory studies (Landau, 1995; Rudman & Glick, 1999; Sinclair & Kunda, 1999). Given the persistence of prejudice and discrimination, it is surprising that little research has been conducted on prejudice reduction since early investigations on eliminating traditional prejudice (e.g., Rokeach's value confrontation procedure). Over the years, the nature of prejudice has changed, such that modern forms of prejudice (e.g., aversive racism) have emerged (Gaertner & Dovidio, 1986). Thus, it is sensible to presume that new methods of prejudice reduction are needed (Monteith, Zuwerink, & Devine, 1994).

Recent research (Aronson, 1999) has demonstrated that, under certain conditions, hypocrisy induction is a successful social influence technique. Thus, the goal of the current study was to use hypocrisy induction to reduce a modern form of prejudicial or discriminatory behavior. In particular, we hypothesize that a hypocrisy induction should successfully reduce discriminatory behavior that stems from aversive racism. According to Gaertner and Dovidio (1986), aversive racists consciously endorse nonprejudicial egalitarian attitudes but subconsciously have negative feelings toward Blacks because of the history of racism in their culture and because of cognitive processing biases (e.g., in-group bias). Aversive racists experience themselves to be nonprejudiced yet have unacknowledged negative attitudes toward out-group members. Interestingly, the conflicting prejudicial attitudes that aversive racists hold should make them particularly susceptible to a hypocrisy induction. Let us elaborate.

For a hypocrisy experience to initiate behavior change, first, an individual must publicly advocate an attitude-consistent behavior. For individuals to feel hypocritical regarding prejudice, they must have nonprejudicial standards that are internalized and that can be expressed. Aversive racists strive to be egalitarian and truly believe that prejudice and discrimination are wrong (Gaertner & Dovidio, 1986). Thus, aversive racists should have nonprejudicial standards that they could publicly advocate.

Second, the individual must recognize past instances in

This project was funded by a Social Sciences and Humanities Research Council of Canada (SSHRC) doctoral fellowship awarded to the first author and an SSHRC research grant awarded to the third author. An earlier version of this report was presented in poster format at the first annual meeting of the Society of Personality and Social Psychology (Son Hing, Li, & Zanna, 2000). We thank Patricia Devine for lending us her affect measure and Kristen Vance and Patricia Devine for lending us their hypocrisy induction procedure and materials (which we adapted). We also thank Ramona Bobocel, Patricia Devine, Steve Spencer, Timothy Wilson, and anonymous reviewers for their helpful comments and suggestions on an earlier draft of this report. Finally, we thank the research assistants involved with this project: Jeff Bennett, Rick Byers, and Wendy Wood.

Address correspondence and reprint requests to Leanne S. Son Hing, Department of Psychology, University of Guelph, Guelph, Ontario, Canada N1G 2W1. E-mail: sonhing@psy.uoguelph.ca.

0022-1031/01 $35.00

which he or she, personally, did not uphold the advocated standard. Despite aversive racists' nonprejudicial personal standards, we suspect that they do not always practice what they preach. According to Gaertner and Dovidio (1986), aversive racists will discriminate against out-group members in situations for which appropriate intergroup behavior is ambiguous or for which a non-race-related justification exists for discrimination. Furthermore, it has been found that people of all prejudice levels imagine that their actual reactions toward out-group members would be more prejudiced than they should be (Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1996). Although aversive racists supposedly rationalize discriminatory behavior with non-race-related justifications (Gaertner & Dovidio, 1986), we assume that they are able to recall instances in which they had negative responses to out-group members.

Third, for a hypocrisy experience to initiate behavior change, the speaker must be presented with the opportunity to change behavior in a way that allows him or her to practice what was just preached. After facing their own hypocrisy regarding prejudicial responses, aversive racists should be strongly motivated to behave in a less prejudiced manner than is typical. Usually, however, aversive racists are successful at suppressing their prejudicial feelings:

> When a situation or event threatens to make the negative portion of their attitude salient, aversive racists are motivated to repudiate or dissociate these feelings from their self-image, and they vigorously try to avoid acting wrongly on the basis of these feelings. In these situations, aversive racists may overreact and amplify their positive behavior in ways that would reaffirm their egalitarian convictions and their apparently nonracist attitudes. (Gaertner & Dovidio, 1986, p. 62)

Consequently, Dovidio and Gaertner (1991) suggested that increasing aversive racists' awareness of their subtle prejudiced responses should motivate behavior change because of the inconsistency with their egalitarian attitudes. To date, however, no study has empirically tested methods of prejudice reduction among aversive racists (Monteith et al., 1994).

So why might experiences of hypocrisy motivate behavior change among aversive racists? Devine and Monteith's work on prejudice with compunction revealed that when people hold strong nonprejudicial standards, imagining or experiencing strong discrepancies between their "should" and "would" prejudicial responses (Devine et al., 1991; Monteith, 1996) creates feelings of guilt and discomfort. Because a hypocrisy induction should make salient the discrepancies between aversive racists' nonprejudicial ideals and their prejudicial slip-ups, we hypothesize that a hypocrisy induction should evoke negative feelings (e.g., guilt, discomfort) and reduce prejudicial behavior in aversive racists. However, in contrast to Devine and Monteith's work on prejudice with compunction, we do not believe that such an effect would occur for simply everyone.

Specifically, a hypocrisy induction procedure should not reduce discriminatory behavior for people who are truly low in prejudice. It is possible that people who are low in prejudice have no—or only minor—prejudicial slips that can be recalled. Subsequently, to recall a prejudicial slip might not induce negative feelings or motivate a change in behavior for low prejudiced people. Therefore, in the current study, we compared the effects of a hypocrisy induction procedure on people who, we have reason to believe, are truly low in prejudice versus people who are aversive racists.

To identify aversive racists, we created a measure of aversive racism because currently none existed. According to Gaertner and Dovidio (1986), aversive racists are characterized by their ambivalent attitudes toward out-group members; they hold egalitarian values as well as unacknowledged negative feelings and beliefs. Influenced by recent work on automatic and controlled processes and on implicit and explicit attitudes (e.g., Greenwald & Banaji, 1995), we reconceptualized aversive racists as individuals with a discrepancy between their prejudicial attitudes at the implicit versus explicit level (see also Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Wilson, Lindsey, & Schooler, 2000). To our knowledge, this is one of the first times that the discrepancy between implicit and explicit measures has been used to measure a construct or to identify a group of individuals.[1]

Based on our reading of aversive racism theory (Gaertner & Dovidio, 1986), we presume that aversive racists have negative automatic evaluations of out-group members of which they typically are unaware. Thus, aversive racists can be conceptualized as high in implicit prejudice. However, aversive racists have internalized nonprejudiced and egalitarian standards at a conscious level, and so they can be conceptualized as low in explicit prejudice. Therefore, we identified aversive racists as those who are high in implicit prejudice but low in explicit prejudice, and we identified truly low prejudiced individuals as those who are low in both implicit and explicit prejudice. Consistent with our expectations for low prejudiced people, it has been found that people low in implicit prejudice rarely exhibit even subtle forms of discrimination (Dovidio et al., 1997; Fazio, Jackson, Dunton, & Williams, 1995).

Theoretically, a discrepancy between people's implicit and explicit prejudice is possible because measures of implicit and explicit prejudice are found to be weakly related (Dovidio et al., 1997; Fazio et al., 1995; but see Wittenbrink, Judd, & Park, 1997). This is because people's motivation to control prejudice, specifically their concern with acting prejudice, modifies the relation between expressions of implicit and explicit prejudice (Fazio et al., 1995). Of consequence to the identification of aversive racists, there is evidence that people who are high in implicit prejudice but low in explicit prejudice tend to be high in motivation to

[1] Recently, Dovidio (2000) independently seemed to come to the same conclusion on how to identify aversive racists.

TABLE 1
Asian Modern Racism, External Motivation to Control Prejudicial Responses, and Implicit Prejudice

Scale	Intercorrelations AMRS	Intercorrelations EMS	Median	*M*	*SD*	Minimum score	Maximum score
AMRS	(.71)		–0.89	–0.87	0.97	–3.56	2.00
EMS	.17**	(.85)	5.00	4.89	1.72	1.00	9.00
Implicit	–.02	.12	0.00	–0.49	1.00	–2.00	2.00

Note. Cronbach's alpha estimates of internal consistency are in parentheses. AMRS, Asian Modern Racism Scale ($N = 284$); EMS, External Motivation Scale ($N = 275$ because of missing data); Implicit, implicit prejudice ($N = 49$).
** $p < .01$.

control prejudice (Dunton & Fazio, 1997; Fazio et al., 1995). Theoretically, aversive racists should be motivated to be nonprejudiced because they have internalized such standards and not because they experience societal pressure to be politically correct. Thus, we identified aversive racists as individuals who score low on both explicit prejudice *and* external motivation to be nonprejudiced *but* who score high on implicit prejudice. We chose not to identify people with internalized nonprejudicial standards as those who score high on internal motivation to be nonprejudiced because of the high correlations found with explicit measures of prejudice (Plant & Devine, 1998).

For the current study, we investigated prejudice and discrimination against Asians because they constitute the largest visible minority group on our campus and because little research has been conducted on prejudice toward Asians. Our study had a 2 (control vs. hypocrisy condition) × 2 (low vs. high implicit prejudice) factorial design. We hypothesize that a hypocrisy induction procedure should successfully induce negative feelings in aversive racists and consequently lead to a reduction in their discriminatory behavior, compared to aversive racists in a control condition. In contrast, we hypothesize that a hypocrisy induction procedure should have little effect on negative feelings or discriminatory behavior for truly low prejudiced participants, compared to those in a control condition.

METHOD

Subjects and Procedure

Phase 1: Identifying potential participants. In Phase 1, we identified potential participants, that is, those who are low in explicit prejudice and low in external motivation to be nonprejudiced. As part of a larger mass-testing questionnaire, 312 introductory psychology students completed a 9-item Asian Modern Racism Scale (AMRS) that measures negative attitudes toward Asians. We modified the Waterloo Modern Racism Scale (Bobocel, Son Hing, Davey, Stanley, & Zanna, 1998) by targeting Asians as the out-group. A sample item is, "There are too many Asian students being allowed to attend university in Canada" (–4 = *very strongly disagree* to 4 = *very strongly agree*). Also, we measured the degree to which respondents are motivated to be nonprejudiced toward Asians due to external pressures. Students completed Plant and Devine's (1998) 5-item External Motivation to Respond without Prejudice Scale (EMS), which was modified to indicate Asians as the out-group. A sample item is, "I try to act nonprejudiced toward Asian people because of pressure from others" (1 = *strongly disagree* to 9 = *strongly agree*).

Excluding all Asian participants ($N = 292$, remaining), those who fell below the median on the AMRS and the EMS were selected as potential participants (Table 1). A research assistant randomly contacted 49 students (20 men and 29 women) who met the selection criteria to participate in two supposedly unrelated studies for course credit.

Phase 2: Measuring implicit prejudice. Participants were met by a Chinese experimenter with a strong accent who explained that Study 1 was an investigation of how personality affects the free associations made during a word completion task. To activate the concept "Asian" and to create the strongest prime conceivable, *all* participants interacted with the Chinese experimenter for 5 min before she administered the implicit prejudice measure. The Asian experimenter held up cards with word fragments written on them. For each word fragment, participants said aloud the first word that came to their minds (which they wrote down).

To measure implicit prejudice, we modified Gilbert and Hixon's (1991) word fragment completion task. Gilbert and Hixon used this task to measure stereotype activation (i.e., the positive, negative, and neutral automatic associations primed by an Asian experimenter). In contrast, we needed to tap implicit prejudice or the evaluative nature of participants' automatic associations (i.e., positive vs. negative) primed by an Asian experimenter. Thus, we identified a priori five word completions as associated with the Asian stereotype and evaluative in nature. We considered the word completions SLY or SHY, SHORT, and NIP to reflect a negative implicit evaluation of Asians and SMART and POLITE to reflect a positive implicit evaluation of Asians.[2]

[2] Of the participants, 15% made the completion sly, 3% shy, 11% short, 26% nip, 36% smart, and 55% polite.

To create an implicit prejudice scale, positive words (i.e., smart and polite) were reverse keyed and added to the negative word totals, such that higher scores indicate greater implicit prejudice.[3] A median split (median = 0.00) was employed to categorize participants as either low or high implicit prejudice (Table 1).

Phase 3: Hypocrisy induction and subsequent affect. Participants were led to a second, ostensibly unrelated experiment conducted by a Caucasian research assistant who was blind to the participants' level of implicit prejudice. Participants were randomly assigned to the control or hypocrisy condition. The hypocrisy induction procedure was borrowed and adapted from Vance and Devine (1999). Participants were told that Study 2 was an investigation of how personality affects persuasive writing styles. Participants were asked to write persuasive essays on why they believe it is important to treat minority students on campus fairly. Furthermore, participants were told that excerpts from their essays might be featured in pamphlets being dispersed to promote the "Racial Equality Forum," which aimed to enlighten incoming students of the importance of treating other students fairly regardless of race, gender, sexual orientation, and so forth. Ostensibly, participants were randomly chosen to write essays about treating Asian students fairly. Previous research has found that a hypocrisy induction is effective only if people make an initial public declaration of their pro-attitudinal stance (Stone, Aronson, Crain, Winslow, & Fried, 1994). Thus, all participants publicly advocated a nonprejudicial stance toward Asians. After writing their essays, participants in the hypocrisy condition only were instructed as follows:

> The psychology department is interested in understanding more about situations in which Asian students are *not* treated fairly . . . . Please take a few moments and briefly write about two situations in which you reacted more negatively to an Asian person than you thought you should or treated an Asian person in a prejudiced manner.

All participants then completed an affect measure (Monteith, Devine, & Zuwerink, 1993). Participants rated how well each affect word (e.g., guilty) described how they were feeling after writing their essays in the control condition (1 = *does not apply at all* to 7 = *applies very much*) or after writing their essays and describing their personal examples in the hypocrisy condition. We investigated participants' negative feelings of guilt and discomfort with a 12-item composite (e.g., guilty, ashamed, uneasy, uncomfortable) that had a Cronbach's alpha of .94 (Monteith et al., 1993).

Phase 4: Discrimination against Asians. Finally, all participants completed a behavioral measure of discrimination that was adopted and modified from Haddock, Zanna, and Esses (1993). Participants were told that both studies were finished; however, the Caucasian research assistant asked participants to fill out an allegedly anonymous ballot being distributed by the university's Federation of Students (FEDS). The ballot concerned the financial cuts that various student clubs would receive from the FEDS in the coming year. Apparently, the FEDS wanted input from the student body on how to distribute a 20% (or $1,000) cut in funding across 10 student groups. The main dependent variable was the percentage of budget cuts made to the Asian Students' Association's (ASA) budget. Importantly, because the FEDS needed to make budget cuts, a legitimate excuse exists for participants to discriminate against Asians. Thus, we thought that the budget reduction exercise would detect subtle discrimination against Asians. After completing the ballot, participants were probed for suspicion. They were then fully debriefed.

RESULTS

Preliminary Results

Attrition and random assignment. Two participants did not provide complete data; thus, they were excluded from the analyses. Preliminary analyses revealed no main or interactive effects of gender, so we collapsed across gender for all analyses. A 2 (Condition: control vs. hypocrisy) × 2 (Implicit Prejudice: low [low prejudice] vs. high [aversive racist]) analysis of variance (ANOVA) revealed that random assignment was successfully achieved for participants' EMS and implicit prejudice scores but not for their AMRS scores (Table 2). Participants in the hypocrisy condition had marginally greater explicit prejudice scores ($M = -1.48$, $SD = 0.54$), compared to participants in the control condition ($M = -1.83$, $SD = 0.68$). Because there was a relation between AMRS and the main dependent variable (i.e., cuts to the ASA, $r(45) = .25$, $p = .09$), we controlled for participants' level of Asian modern racism in all subsequent analyses.[4]

Participants' essays and examples. Two independent judges, blind to condition and level of implicit prejudice, rated the strength of participants' anti-prejudice essays ("This essay presents a strong/persuasive pro-equality [i.e., anti-racism] viewpoint") and the severity of participants' examples of prejudiced transgressions ("This was a serious/severe example of anti-Asian [i.e., racist] sentiment") on 7-point scales (1 = *strongly disagree* to 7 = *strongly agree*). Participants' two examples were averaged. The judges' ratings for the strength of the essays were averaged, $r(45) = .50$, $p < .001$, as were their ratings for the severity of the examples, $r(21) = .74$, $p < .001$). To assess whether all participants wrote comparably strong essays, we conducted a 2 (Condition) × 2 (Implicit Preju-

[3] When a measure of stereotype activation (i.e., positive, negative, and neutral words [e.g., RICE]) is used instead of implicit prejudice, no significant effects are found in our major analyses. This suggests that when Asians are the target group, implicit stereotyping and implicit prejudice are very different constructs. In fact, they were not related, $r(45) = -.04$, *ns*.

[4] Asian modern racism did not interact with any predictors in any analyses.

TABLE 2
Reactions of Low Prejudiced and Aversive Racist Participants to the Experimental Manipulation

Dependent variable	Condition/Participant group: Control[a] Low prejudice[c] (11)	Control[a] Aversive racist[d] (13)	Hypocrisy[b] Low prejudice[c] (12)	Hypocrisy[b] Aversive racist[d] (11)	Analyses of variance and analyses of covariance: *F*s Condition	Implicit prejudice	Condition × implicit prejudice
EMS	3.69	3.72	3.47	4.08	0.05	0.92	0.74
AMRS	–1.91	–1.76	–1.41	–1.54	3.86†	0.01	0.55
Implicit	–1.36	0.54	–1.33	–0.01	1.93	122.52***	2.53
Essay[e]	4.65	4.14	4.53	4.10	0.05	1.61	0.01
Example[e]			4.83	4.76		0.01	
Negative[e,f]	2.51	2.09	2.80	3.71	5.93*	0.41	3.18†

Note. $N = 47$. Cell *n*s are in parentheses. Where appropriate, covariate-adjusted means are presented. EMS, External Motivation Scale; AMRS, Asian Modern Racism Scale; Implicit, implicit prejudice; Essay, judges' ratings of essay strength; Example, judges' ratings of example severity; Negative, negative feelings.

[a] Essay only.
[b] Essay plus examples.
[c] Low implicit/low explicit prejudice.
[d] High implicit/low explicit prejudice.
[e] Asian modern racism was a covariate.
[f] Self-esteem was a covariate.
† $p < .10$.
* $p \leq .05$.
*** $p < .001$.

dice) analysis of covariance (ANCOVA) with AMRS as the covariate. There was no difference in the essays written by low prejudiced participants and aversive racists (Table 2).

We also tested whether low prejudiced participants and aversive racists provided equally severe examples of prejudicial slip-ups. We had thought it possible that aversive racists might recall more severe prejudicial examples, compared to truly low prejudiced participants. Interestingly, ANCOVA results revealed no effect of implicit prejudice on the severity of participants' prejudicial transgressions (Table 2).

We conducted exploratory analyses to test whether low prejudiced participants recalled examples that occurred further in the past, compared to aversive racists. Two judges, who were blind to participants' implicit prejudice, rated the recency of participants' examples (1 = *childhood,* 2 = *high school,* 3 = *early university,* 4 = *prior term,* 5 = *current term*) if they had an impression of when the events occurred. It was possible to rate the recency of at least one event for 19 participants. The agreement coefficients exceeded .89 for both Examples 1 and 2. Results revealed that aversive racists tended to recall more recent examples ($M = 4.55$, $SD = 0.37$), compared to low prejudiced participants ($M = 3.67$, $SD = 1.20$), $t(17) = 2.00$, $p = .06$. In addition, significantly greater variance was found for the recency of low prejudiced participants' examples, $F(7, 10) = 9.82$, $p = .006$. A nonparametric medians test was reliable, $\chi^2(1, N = 19) = 4.23$, $p = .04$. Only 33% of truly low prejudiced participants were above the overall median in recency (median = 4.50), compared to 80% of aversive racists.

In summary, low prejudiced participants and aversive racists wrote equally convincing essays and recalled equally severe prejudicial slip-ups. However, the slip-ups recalled by low prejudiced participants occurred in the more distant past, whereas those recalled by aversive racists converged on the present.

Affect Results

To investigate participants' affect, we conducted a 2 (Condition) × 2 (Implicit Prejudice) ANCOVA.[5] As can be seen in Table 2, results revealed a main effect of condition such that participants experienced more negative affect in the hypocrisy condition ($M = 3.26$, $SE = 0.27$) than in the control condition ($M = 2.30$, $SE = 0.27$). This effect, however, was moderated by participants' level of implicit prejudice. As predicted, aversive racist and low prejudiced participants had different affective responses to the hypocrisy induction procedure. Simple effects analyses revealed that aversive racists in the hypocrisy condition experienced more negative affect, compared to those in the control condition, $F(1, 39) = 10.53$, $p = .003$. In contrast, among truly low prejudiced participants, experimental con-

[5] AMRS and Rosenberg's (1965) Self-Esteem Scale (from mass testing) were the covariates. The higher participants' self-esteem, the less negative they felt, $F(1, 39) = 6.23$, p = .02. Self-esteem did not interact with any predictors.

dition did not influence participants' negative feelings, $F(1, 39) = 0.18$, *ns*. Thus, supporting our hypothesis, a hypocrisy induction did make aversive racists feel more negative, compared to their counterparts in the control condition; however, it did not affect negative feelings for truly low prejudiced people.

Looking at the interaction from a different perspective, in the control condition, participants' level of implicit prejudice did not influence their negative feelings, $F(1, 39) = 0.96$, *ns*. In contrast, in the hypocrisy condition, there was a significant simple effect of implicit prejudice, $F(1, 39) = 3.86$, $p = .05$, such that aversive racists experienced more negative feelings, compared to low prejudiced participants.

Discrimination Results

To investigate our main dependent variable, participants' discriminatory behavior, we calculated the percentage of cuts made to the ASA. A 2 (Condition) × 2 (Implicit Prejudice) ANCOVA was conducted.[6] Results of the ANCOVA revealed a marginal effect of Asian modern racism, $F(1, 41) = 3.63$, $p = .06$. Participants higher in explicit prejudice tended to make greater cuts to the ASA. There was neither a main effect for condition, $F(1, 41) = 2.59$, *ns*, nor one for implicit prejudice, $F(1, 41) = 0.83$, *ns*. However, there was the predicted Condition × Implicit Prejudice interaction, $F(1, 41) = 9.54$, $p = .004$ (Fig. 1).[7] As hypothesized, aversive racists and low prejudiced participants had different behavioral responses to the hypocrisy induction procedure. Simple effects analyses revealed that experimental condition significantly affected the amount of cuts made to the ASA among aversive racists, $F(1, 39) = 13.40$, $p = .001$, such that those in the hypocrisy condition made fewer cuts to the ASA ($M = 6.5\%$, $SE = 4.21$), compared to those in the control condition ($M = 26.9\%$, $SE = 4.04$). In contrast, among truly low prejudiced participants, experimental condition had no effect on cuts made to the ASA, $F(1, 49) = 1.57$, *ns*, (hypocrisy $M = 23.6\%$, $SE = 4.64$; control $M = 17.5\%$, $SE = 4.47$). Thus, the hypocrisy induction procedure prompted aversive racists, but not low prejudiced people, to reduce their discriminatory behavior toward Asians.

Looking at the interaction in a different way, in the control condition, aversive racist participants tended to make greater cuts to the ASA than did truly low prejudiced participants, $F(1, 41) = 2.90$, $p = .09$.[8] A directional nonparametric test was reliable, $\chi^2(1, N = 24) = 3.23$, $p = .04$. Whereas only 18% of truly low prejudiced participants cut more than 20% to the ASA, 54% of aversive racists made more than the recommended cuts. Thus, when a justification existed, aversive racists were more likely to evince discriminatory behavior, compared to low prejudiced participants.

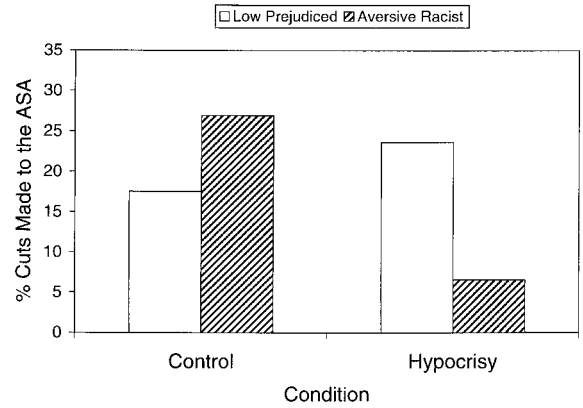

FIG. 1. The condition × implicit prejudice interaction on participants' cuts to the Asian Students Association (ASA). Control, essay-only condition; Hypocrisy, essay plus examples condition; Low Prejudiced, low implicit/low explicit prejudice; Aversive Racist, high implicit/low explicit prejudice.

Simple effects analyses revealed that in the hypocrisy condition, aversive racists made fewer cuts to the ASA, compared to truly low prejudiced participants, $F(1, 41) = 7.60$, $p < .01$. In addition, a directional nonparametric test was reliable, $\chi^2(1, N = 23) = 5.79$, $p = .008$. Whereas 45% of truly low prejudiced participants cut 20% or less to the ASA, 92% of aversive racists made fewer cuts to the ASA than were required. Surprisingly, 42% of aversive racist participants either did not cut funds or gave additional funding to the ASA (no low prejudiced people failed to make cuts to the ASA). Thus, when aversive racists were made aware of their prejudicial tendencies, they "bent over backward" to act in a nonprejudicial manner.

DISCUSSION

Results confirmed our hypothesis that a hypocrisy induction should result in negative feelings (e.g., guilt, discomfort), and subsequently a reduction in prejudicial behavior,

[6] AMRS and the average amount of money given to the nine other student groups were the covariates. Only 25 participants accurately proportioned funds such that the budget total reached $4,000. To take into account the tendency to under- or overcut funding, we controlled for the average amount of cuts that participants made to the nine other groups, although the same pattern of interaction is found when AMRS is the only covariate used, $F(1, 42) = 9.16$, $p = .004$.

[7] When multiple regression analysis is conducted with implicit prejudice as a continuous variable, a significant Condition × Implicit Prejudice interaction emerges, $B = -5.31$, $p = .02$, with a significant simple effect of implicit prejudice in the hypocrisy condition, $B = -7.79$, $p = .03$, and a significant simple effect of condition among aversive racists, $B = -12.17$, $p = .003$). The two other simple effects were not significant.

[8] Nine participants indicated some suspicion that the FEDS ballot might have been part of the study; however, none indicated that it was intended to measure discrimination against Asians. Interestingly, when their data were excluded, our results became stronger. In particular, in the control condition, aversive racists made greater cuts to the ASA ($M = 30.5\%$, $SE = 4.93$), compared to low prejudiced participants ($M = 18.2\%$, $SE = 4.81$), $F(1, 32) = 3.85$, $p = .06$.

for aversive racists but not for participants truly low in prejudice. Interestingly, low prejudiced and aversive racist participants wrote equally strong essays advocating the importance of being nonprejudiced, and in the hypocrisy condition they recalled equally severe examples of prejudicial slip-ups, although exploratory analyses revealed that truly low prejudiced participants recalled slip-ups that occurred in the more distant past. Subsequently, aversive racists in the hypocrisy condition experienced more negative feelings, compared to those in the control condition. Finally, a hypocrisy induction motivated aversive racists to correct for their transgressions by reducing cuts to the ASA or, in some instances, by giving the ASA additional funds. We believe that the hypocrisy induction procedure forced aversive racists to become aware of the negative aspects of their attitudes that they typically repress (Wilson et al., 2000). An interesting issue to address in future research is the degree to which aversive racists are typically aware of their implicit prejudice (T. Wilson, personal communication, April 19, 2000).

Truly low prejudiced participants had a different reaction to the hypocrisy induction: They did not feel more negative and did not cut funds to the ASA less, compared to those in the control condition. It is possible that low prejudiced participants had different affective and behavioral responses to the hypocrisy induction than did aversive racists because the former had to search back in time to recall events in which they had prejudicial responses to Asians. Possibly, truly low prejudiced participants experienced more difficulty in retrieving examples of prejudicial slip-ups and interpreted such difficulty as a sign that they rarely are hypocritical (Schwarz et al., 1991). Also, it is possible that low prejudiced participants disassociated their prejudicial slip-ups from their current self-images because the events had occurred sufficiently in the past (Ross & Wilson, 2000). Both of these processes could be investigated in future research.

The results we obtained provide construct validity for our reconceptualization and measurement of aversive racism. Thus, we can tentatively conclude that it is possible to classify aversive racists as individuals low in explicit prejudice but high in implicit prejudice. Participants classified as aversive racists behaved in a manner entirely consistent with aversive racism theory (Gaertner & Dovidio, 1986). First, given that a prejudicial response could be justified in the control condition, aversive racists tended to discriminate against Asians—even after advocating nonprejudice. It would be interesting to test the discriminatory behavior of aversive racists without such a reminder of their egalitarian standards. Second, in the hypocrisy condition, aversive racists bent over backward to treat Asians positively when their negative attitudes were made salient.

It appears that aversive racists needed to experience consciousness raising and subsequent negative feelings to avoid behaving in a discriminatory manner. These results support Monteith's (1993) model of the self-regulation of prejudiced responses. Monteith hypothesized that, when people are made aware of the discrepancy between their ideal and actual prejudicial responses, negative self-directed affect acts as a punishment that motivates nondiscriminatory future behavior. Interestingly, the results of the current study suggest that such processes operate for aversive racists but not for truly low prejudiced people (see also Monteith & Voils, 1998).

One implication of the current research is that the classic aversive racism studies (e.g., Gaertner & Dovidio, 1977) could be rerun with this individual difference measure of aversive racism. Presumably, participants low in both implicit and explicit prejudice rarely, if ever, should discriminate against out-group members. In contrast, participants high in implicit prejudice but low in explicit prejudice should discriminate but only in ambiguous situations. Future research should also investigate aversive racism using different implicit prejudice measures and different targets of discrimination.

Given that a hypocrisy induction procedure can successfully reduce prejudicial behavior for aversive racists, an important question arises: Might such experiences also serve to reduce prejudicial behavior for other groups? In addition to truly low prejudiced people, a hypocrisy induction might not work for bigots (i.e., people high in explicit prejudice) because they tend to have relatively low and weakly internalized nonprejudicial standards (Devine et al., 1991). Furthermore, when bigots imagine violating these standards, rather than feeling negative about themselves, they feel negative toward others (Monteith et al., 1993). Testing a hypocrisy induction procedure on people high in explicit prejudice, however, would help to explicate the current findings. In addition, a hypocrisy manipulation might be unsuccessful among those whose standards of nonprejudiced behavior stem from societal pressures (e.g., to be politically correct). A hypocrisy induction might induce feelings of fear and threat (Higgins, 1987) and consequently increase tendencies to stereotype and discriminate (Fein & Spencer, 1997).

A second important question that arises is whether hypocrisy induction might help to actually reduce implicit prejudice. Recent work by Kawakami, Dovidio, Moll, Hermsen, and Russin (2000) provides intriguing evidence that practice at stereotype negation can reduce automatic stereotyping. Through a similar process, it is possible that aversive racists can develop positive automatic evaluations of out-group members. Importantly, experiences of hypocrisy, consciousness raising, and feelings of guilt and discomfort might motivate aversive racists to reduce their negative automatic reactions toward out-group members. However, it is also possible that aversive racists have failed to develop positive automatic evaluations of out-groups because of the emotional beating they give themselves on the rare occasions when they acknowledge their prejudice.

In other words, strong negative affect might disrupt the learning of new positive automatic associations. In future research, it will be important to look at the processes by which aversive racists may develop positive implicit attitudes and the role of affect in implicit attitude change more generally.

REFERENCES

Aronson, E. (1999). Dissonance, hypocrisy, and the self concept. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 103–126). Washington, DC: American Psychological Association.

Bobocel, D. R., Son Hing, L. S., Davey, L. M., Stanley, D. J., & Zanna, M. P. (1998). Justice-based opposition to social policies: Is it genuine? *Journal of Personality and Social Psychology,* **75,** 653–669.

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

Dovidio, J. F. (2000, June). *The psychology of prejudice: The third wave.* Presidential address delivered at the meeting of the Society for the Psychological Study of Social Issues, Minneapolis, MN.

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

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

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

Fazio, R. H., 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.

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology,* **73,** 31–44.

Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of White racism, arousal, and helping behavior. *Journal of Personality and Social Psychology,* **35,** 691–707.

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. 61–86). Orlando, FL: Academic Press.

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

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

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology,* **65,** 1105–1118.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review,* **94,** 319–340.

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.

Landau, J. (1995). The relationship of race and gender to managers' ratings of promotion potential. *Journal of Organizational Behavior,* **16,** 391–400.

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

Monteith, M. J. (1996). Affective reactions to prejudice-related discrepant responses: The impact of standard salience. *Personality and Social Psychology Bulletin,* **22,** 48–59.

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

Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudicial responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology,* **75,** 901–916.

Monteith, M. J., Zuwerink, J. R., & Devine, P. G. (1994). Prejudice and prejudice reduction: Classic challenges, contemporary approaches. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology* (pp. 323–346). Toronto: Academic Press.

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

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton Univ. Press.

Ross, M., & Wilson, A. E. (2000). Constructing and appraising past selves. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 231–258). Cambridge, MA: Harvard Univ. Press.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women. The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology,* **77,** 1004–1010.

Schwarz, N., Bless, H., Strack, F., Klumpp, G. K., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology,* **61,** 195–202.

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.

Son Hing, L. S., Li, W., & Zanna, M. P. (2000, February). *Can hypocrisy reduce prejudice, especially for those low in explicit but high in implicit prejudice?* Poster presented at the annual meeting of the Society of Personality and Social Psychology, Nashville, TN.

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin,* **20,** 116–128.

Vance, K. M., & Devine, P. G. (1999, May). *How self-affirmations influence the motivation and behaviors associated with prejudice reduction.* Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.

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.