
Types of High Self-Esteem and Prejudice: How Implicit Self-Esteem Relates to Ethnic Discrimination Among High Explicit Self-Esteem Individuals

Christian H. Jordan
Steven J. Spencer
Mark P. Zanna
University of Waterloo

There is increasing recognition that high self-esteem is heterogeneous. Recent research suggests that individuals who report having high self-esteem (i.e., have high explicit self-esteem) behave more defensively to the extent that they have relatively low implicit self-esteem. The current studies test whether individuals with high explicit self-esteem are more likely to discriminate ethnically, as a defensive technique, to the extent that they have relatively low implicit self-esteem. The results support this prediction. Among participants with high explicit self-esteem, all of whom were threatened by negative performance feedback, those with relatively low implicit self-esteem recommended a more severe punishment for a Native, but not a White, student who started a fist-fight. In Study 2, this pattern was not apparent for participants with relatively low explicit self-esteem.

Keywords: *self-esteem; implicit; defensive; prejudice; discrimination*

Recently, Carla Bozulich, a young, relatively unknown musician, received a rare opportunity. Willie Nelson, a respected pioneer of country music, offered to play with her to contribute to an avant-garde album she was recording covering his songs. After the recording sessions, when asked what Nelson was like, Bozulich replied, "Incredibly gracious. Confident in a way that doesn't need to take from others to be strong" (Keenan, 2003, p. 14). This impression may be revealing not simply of Nelson's personality but also of two basic insights about self-esteem and interpersonal relations. First, Bozulich's remarks suggest that self-confidence, or self-esteem, can assume different forms. Although Nelson appeared gracious in his confidence, by implication, others might be more selfish. The second, related insight is that people who have a less gracious form of self-esteem may often put

other people down or "take from others" in order to bolster their own positive self-views.

There is now considerable evidence that high self-esteem is heterogeneous. Kernis (2003; Kernis & Paradise, 2002) reviewed numerous studies suggesting that some high self-esteem people are relatively secure in their positive self-views, whereas others have fragile self-views that require continual reinforcement (see also Deci & Ryan, 1995; Epstein & Morling, 1995). Moreover, one conclusion of a recent monograph exploring the potential benefits of high self-esteem (and costs of low self-esteem) was that high self-esteem per se is not a fine-grained enough distinction; there is further need to distinguish between different forms that high self-esteem can assume (Baumeister, Campbell, Krueger, & Vohs, 2003). Our own past research (Jordan, Spencer, & Zanna, 2003; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003) indicates that individuals with high self-esteem may be relatively secure or defensive, depending on their levels of implicit self-esteem.

Implicit self-esteem refers to highly efficient self-evaluations that may exist largely outside of awareness (Epstein & Morling, 1995; Greenwald & Banaji, 1995).

Authors' Note: This research was supported by a doctoral fellowship from the Natural Sciences and Engineering Research Council of Canada to the first author and by grants from the Social Sciences and Humanities Research Council of Canada to the second and third authors. We are grateful to Gaylord Albrecht and Jeff Spence for their assistance with data collection. Correspondence concerning this article should be addressed to Christian Jordan, now at the Department of Psychology, Wilfrid Laurier University, 75 University Avenue West, Waterloo, Ontario, Canada, N2L 3C5; e-mail: cjordan@wlu.ca.

PSPB, Vol. 31 No. 5, May 2005 693-702
DOI: 10.1177/0146167204271580

© 2005 by the Society for Personality and Social Psychology, Inc.

Self-esteem as traditionally conceptualized, in contrast, can be considered to be *explicit*; that is, deliberately reasoned and controlled. Whereas explicit self-esteem is measured by standard self-report scales, implicit self-esteem is measured by less direct means, often by computer-based reaction time tasks that require responses that are difficult to control (Bosson, Swann, & Pennebaker, 2000; Jordan, Spencer, & Zanna, 2003). Interestingly, measures of implicit and explicit self-esteem are largely independent. Knowing that an individual reports having high explicit self-esteem reveals virtually nothing about his or her implicit self-esteem, which could be quite high or quite low. Some people may thus possess two distinct self-evaluations with different valences, one explicit and the other implicit (cf. Smith & DeCoster, 2000; Wilson, Lindsey, & Schooler, 2000).

A number of classic perspectives posit that individuals who consciously feel positively about themselves, but who harbor self-doubts and insecurities at less conscious levels, may be particularly defensive (e.g., Coopersmith, 1959; Harder, 1984; Horney, 1937). Until recently, however, this possibility remained largely untested. Noting the relative independence of measures of explicit and implicit self-esteem, we used such measures to operationalize and test these classic perspectives (Jordan, Spencer, & Zanna, 2003; Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). We thus predicted and found that individuals with high explicit but low implicit self-esteem behave more defensively than do individuals who are high in both explicit and implicit self-esteem. Specifically, among individuals with high explicit self-esteem, we found that those with relatively low implicit self-esteem rationalized their decisions more and favored in-group members more in a minimal group context (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003).

These findings may help explain why high explicit self-esteem is sometimes associated with behaviors that can have negative social repercussions. When people with positive self-views are challenged by criticism or negative feedback, they sometimes react in ways that disadvantage other people. For instance, individuals with relatively positive self-views sometimes behave more aggressively toward others (Bushman & Baumeister, 1998); interact in less genial, more antagonistic ways (Heatherton & Vohs, 2000); and discriminate more (Crocker, Thompson, McGraw, & Ingerman, 1987). Each of these tendencies can be viewed as a means of defending threatened self-views. Put simply, when people feel poorly about themselves, they may sometimes derogate or disadvantage others in order to feel better (e.g., Wills, 1981). Prejudice, moreover, may be a special instance of this process.

Wills (1981) suggested that prejudice is a form of downward social comparison aimed at enhancing subjective well-being and protecting self-esteem. Similarly, a key premise of Tajfel and Turner's (1979) *social identity theory* is that people often favor in-group members over out-group members in order to enhance their own self-worth. One corollary of this premise is that people should discriminate more when their self-worth is acutely low or threatened (Hogg & Abrams, 1990; Wills, 1981). Fein and Spencer (1997) found that people who had their positive self-views threatened by negative performance feedback were indeed more likely to use negative stereotypes to judge out-group members. To the extent that they did so, moreover, they felt better about themselves. It thus appears that discrimination can, at least sometimes, be used as an effective self-esteem maintenance technique.

A second corollary of social identity theory, which has received less empirical support, is that people with chronically low self-esteem ought to discriminate more (Hogg & Abrams, 1990; see also Wills, 1981). In fact, the available evidence suggests that people with high self-esteem are most prone to engage in in-group bias in minimal group situations (Aberson, Healy, & Romero, 2000). They also, in some cases, discriminate more against existing out-groups (Crocker et al., 1987) and express more racist attitudes (Emler, 2001). It may thus be individuals with high explicit self-esteem who discriminate most to protect their self-views. Nevertheless, we believe it is primarily individuals with high explicit but low implicit self-esteem who discriminate in this way. This prediction is consistent with our earlier finding that individuals with high explicit but low implicit self-esteem engage in more in-group bias in the minimal group paradigm than do individuals with high explicit and implicit self-esteem (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). The minimal group paradigm was, however, designed specifically to rule out preexisting histories of prejudice between groups as an explanation for bias (Tajfel & Turner, 1979). As a consequence, we cannot conclude that explicit and implicit self-esteem are similarly related to discrimination between extant social groups.

STUDY 1

In Study 1, we explored whether people with high explicit self-esteem differ in their tendencies to discriminate against members of existing out-groups when their self-views are challenged. To do so, we had participants (all of whom were preselected for having high explicit self-esteem) complete the Implicit Associations Test (IAT) as a measure of implicit self-esteem. We then had them complete a bogus intelligence test, for which they received negative performance feedback, in order to

threaten their positive self-views (following Fein & Spencer, 1997). Participants next read a series of cases of student misconduct. In one case, a student is asleep on a couch outside a campus bar late at night. He is disturbed by a group of students exiting the bar and exchanges words with them. The situation escalates until the offender is insulted and retaliates with physical aggression. We manipulated the apparent ethnicity of the student offender in this case such that he appeared to be either Native or White. We chose to focus on Native Canadians as the ethnic out-group because attitudes toward Natives are often relatively negative. When independent samples of University of Waterloo students were asked to indicate their attitudes toward several different ethnic groups, they expressed the most negative attitudes toward Natives (rivaled only by attitudes toward Arabs and Pakistanis; Esses, Haddock, & Zanna, 1993; Esses & Zanna, 1995; Haddock & Zanna, 1994) and expressed negative stereotypes about Natives, including that they are alcoholic, lazy, poor, and uneducated (Haddock & Zanna, 1994). After reading the case, participants indicated the severity of punishment they believed to be appropriate for the offender. We expected our high explicit self-esteem participants to recommend a more severe punishment for a Native, but not a White, student offender to the extent that they had relatively low implicit self-esteem. This would suggest that individuals with high explicit but low implicit self-esteem are more likely than those high in both explicit and implicit self-esteem to discriminate as a means of defending their self-views.

Method

PARTICIPANTS

Thirty-seven students enrolled in introductory psychology at the University of Waterloo participated in exchange for partial course credit. Only participants identified as having high explicit self-esteem, as described below, were recruited for the study. Two participants' data were excluded from analyses because their error rates on the IAT exceeded 20% (Greenwald, McGhee, & Schwartz, 1998). This left 35 participants (22 female). Most participants were White (88.5%); the remaining participants were Asian.

MATERIALS AND PROCEDURE

Preselecting individuals with high explicit self-esteem. All students enrolled in introductory psychology at the University of Waterloo during one academic term completed the Rosenberg (1965) Self-Esteem Scale (RSES) as part of a mass pretesting booklet. The RSES contains 10 items, such as "I feel I am a person of worth, at least on an equal basis with others." Respondents indicated their level of agreement with each item on a 9-point scale

ranging from (1) *very strongly disagree* to (9) *very strongly agree*. Items are scored such that higher numbers reflect higher self-esteem. Respondents who scored in the top third of the distribution (7.5 or higher) were invited to participate.

Participants were scheduled in groups of one to four. They were met by a male experimenter who described the purpose of the study as an investigation of intellectual abilities and judgment styles. They were told that they would complete a series of different tests and tasks that would allow us to examine the relations between perceptual styles, intellectual abilities, and how people make social decisions. Participants signed consent forms and were seated at separate cubicles containing Macintosh computers.

Implicit self-esteem. Participants next completed the IAT measure of implicit self-esteem, which was described as a measure of perceptual style to disguise its relation to self-evaluation. Participants had to categorize words appearing on their computer screens as quickly and accurately as possible. The words to be categorized appeared in the center of the screen, and the category labels appeared in the upper left and right sides of the screen. Participants used the A and K keys to indicate the category to which each word belonged and made two types of categorizations: (a) between 10 pleasant and 10 unpleasant words (i.e., *holiday, warmth, friend, smile, sunshine, gift, love, party, happy, joy, agony, death, disease, vomit, evil, cockroach, pain, stink, disaster, garbage*) and (b) between 2 self and 2 not-self words (i.e., *me, myself, it, that*). During blocks including the pleasant versus unpleasant distinction, each pleasant and each unpleasant word was presented once. During blocks including the self versus not-self distinction, each self and each not-self word was presented five times on separate trials. Within each block of trials, words were presented in random order. There were five blocks of trials in all.

Blocks 1 and 2 were practice blocks for which participants categorized pleasant versus unpleasant words (Block 1) and self versus not-self words (Block 2). Block 3 then combined these categorizations such that participants discriminated unpleasant versus pleasant words and self versus not-self words on separate trials within the same block, using one response key to indicate if a word belonged to the unpleasant or self categories and the other key if the word belonged to the pleasant or not-self categories. During Block 4, participants again practiced self versus not-self discriminations with the response keys now reversed from those they had earlier practiced. Block 5 again combined the two types of categorization on separate trials, with participants now using one response key to indicate if a word belonged to the unpleasant or not-self categories and the other key if the word belonged to the pleasant or self categories.

Data from Blocks 3 and 5 were used to compute IAT scores. The IAT is premised on the logic that individuals with relatively high implicit self-esteem will respond faster when self and pleasant share a single response (Block 5) than when self and unpleasant share a single response (Block 3), because their automatic associations between the self and positive affect will interfere with their responses in the latter but not the former case (Greenwald et al., 1998). IAT scores were thus computed by subtracting participants' average response latencies during Block 5 from their average response latencies during Block 3, with higher scores reflecting higher implicit self-esteem. Before doing so, however, all error trials were excluded. To control the influence of outliers, response latencies longer than 3,000 ms were recoded as 3,000 ms, and response latencies shorter than 300 ms were recoded as 300 ms (Greenwald et al., 1998). As in Jordan, Spencer, Zanna, Hoshino-Browne, et al. (2003), the not-self words in the IAT were chosen to be as neutral as possible (i.e., *it*, *that*) rather than reflecting the concept of "other" (e.g., *they*, *them*; as suggested by Farnham, Greenwald, & Banaji, 1999) in order to avoid confounding affect associated with the self and affect associated with others in IAT scores (see Karpinski, 2004). This seemed particularly desirable in the present context of examining discrimination toward out-groups to ensure that any observed effects reflect self-feelings rather than implicit prejudice.

Negative performance feedback. After finishing the IAT, all participants completed a bogus intelligence test administered by computer, for which they received negative performance feedback. Following Fein and Spencer (1997), participants were told that the test "measures general verbal intelligence and reasoning ability, and has been found to be a good predictor of academic and career success." Further instructions were delivered by computer. Each participant typed in his or her name and identified demographic information. The test was titled "The Reasoning and Verbal Acuity Battery" and was purportedly validated on numerous samples in the United States and Canada. It consisted of five parts, including analogies, antonyms, sentence completions, syllogisms, and verbal-nonverbal matching. This combination of tasks was alleged to be optimal for assessing general intelligence. A description of each task, and its relevance to verbal intelligence and reasoning, was given prior to each part. Each part contained 6 to 10 multiple-choice questions, derived from advanced tests used for admissions to graduate or law school. Some questions were changed such that correct answers were not presented as response options. In addition, the questions had short time limits, and a clock appeared in the corner of the screen counting down the time remaining for each

question. The test was thus designed to be quite difficult, making the negative feedback more plausible.

Once participants had completed the test, the program indicated that it was calculating scores. A few seconds later, the computer presented percentile ranks—relative to other students in the United States and Canada who had completed the test—for each part. Each participant received the same scores ranging from the 33rd to the 56th percentile. In previous research, this same feedback effectively challenged university students' self-views (Fein & Spencer, 1997).

Target race manipulation. Next, participants were asked to complete a social decision-making task. They received booklets containing three vignettes, each describing a case of student misconduct that was ostensibly based on an actual case that appeared before the university's disciplinary committee. Participants were asked to read each case as though they were members of the disciplinary committee who needed to determine an appropriate punishment for the student in question. In each case, a student was clearly guilty of a punishable offence. In the first case, a student damaged a vending machine. In the second, a student was caught smoking marijuana in his dorm room. In the final case, a student started a fistfight outside a campus bar. Participants read one of two versions of this final case, determined by random assignment. Roughly half of the participants read about a student offender named John Proudfoot, whereas the remaining half read about John Pride. This manipulation was intended to convey, respectively, that the student offender was or was not Native in ethnicity. The experimenter was blind to experimental condition up until the point at which participants received the final questionnaire and was blind to their levels of implicit self-esteem throughout the session.

In this final vignette, the student offender (Proudfoot or Pride) is initially asleep on a couch outside a campus bar late on a Friday night, and it is unclear whether he has been drinking. He is disturbed by a group of students leaving the bar and exchanges words with them. A student in the group, "Ian Merritt," insults him by calling him an "asshole," at which point the student offender punches Merritt twice in the face. Hence, the student offender is clearly culpable of physical aggression but has also been clearly provoked.

Severity of punishment measures. After reading each case, participants were asked to indicate what they believed would be an appropriate punishment for the student offender. For the final case, participants first responded to the question, "Student offenders can be required to take anger management courses. To what extent do you agree or disagree that [Proudfoot/Pride] should be required to take an anger management course?"

Participants responded using a 9-point Likert-type scale with end points labeled 1 (*strongly disagree*) and 9 (*strongly agree*). Next, participants were asked, "How severe a punishment does [Proudfoot/Pride] deserve?" They responded using a 9-point scale with end points labeled 1 (*very mild*) and 9 (*very severe*).

Finally, participants selected one punishment from a list of objective possible punishments based on penalties actually employed by the university. The punishments were presented in order of escalating severity, ranging from *no disciplinary action*, to *reprimand*, *disciplinary probation* (of 1 term, 2 terms, 1 year, or 2 years), *suspension* (of 1 term, 2 terms, 1 year, or 2 years), and finally *expulsion*. Each punishment was defined for participants. Thus, for disciplinary probation it was stated that

the student is informed that any further offence while the student is on probation will lead to a significantly stiffer penalty. A note that the student has been placed on probation shall be placed in the student's file, but will not appear on grade reports and transcripts: the note will be removed when the student completes his/her probation period.

After reading the punishments and descriptions, participants indicated the punishment they believed to be most appropriate for the student offender. A response of no disciplinary action was scored as 1, reprimand as 2, a one-term disciplinary probation as 3, and so on, up to 11 for expulsion.

After completing this questionnaire, all participants were thanked and debriefed with a process debriefing (Ross, Lepper, & Hubbard, 1975). Particular attention was given to delineating the deception used in the study. Before leaving the lab, all participants had to demonstrate a clear understanding that their feedback on the intelligence test was false and not reflective of their true abilities.

Results and Discussion

The three items assessing the recommended severity of punishment—recommending the student take an anger management course, and the subjective and objective severity of punishment—correlated significantly with one another and showed reasonable internal consistency (standardized item $\alpha = .68$). We thus converted each item to a z-score form and aggregated the items into a single measure of severity of punishment.

Recall that we expected participants, all of whom had high explicit self-esteem (and all of whom were under ego threat), to discriminate more to the extent that their levels of implicit self-esteem were relatively low. We tested this prediction by regressing IAT scores, target ethnicity (dummy coded; White = 0, Native = 1), and the interaction between the two onto recommended severity

of punishment ratings, after first centering IAT scores. This analysis revealed a significant main effect of implicit self-esteem, $t(32) = -3.55, p = .001$, indicating that participants with lower IAT scores recommended more severe punishments overall ($\beta = -.53$). As expected, however, this main effect was qualified by a significant two-way interaction, $t(31) = -2.09, p = .04$ ($\beta = -.40$). IAT scores were not significantly related to the severity of punishment recommended for John Pride, $t(31) = -1.35, ns$ ($\beta = -.19; y'_{low} = .02, y'_{high} = -.09$).¹ In contrast, there was a significant negative relation between implicit self-esteem and the severity of punishment recommended for John Proudfoot, $t(31) = -4.06, p < .001$ ($\beta = -.58; y'_{low} = .56, y'_{high} = -.64$). Similarly, the interaction indicates that whereas participants with relatively low implicit self-esteem ($-1 SD$) tended to recommend a more severe punishment for John Proudfoot than John Pride (a pattern consistent with racial discrimination), $t(31) = 1.71, p = .09$, participants with higher implicit self-esteem ($+1 SD$) recommended more equivalent punishments, $t(31) = -1.27, ns$.

Thus, consistent with our predictions, among individuals with high explicit self-esteem under threat, those with relatively low implicit self-esteem recommended a more severe punishment for a Native, but not a White, student offender. The full magnitude of this effect is clear when the objective punishment recommendations are considered separately. Participants with high explicit self-esteem who had relatively low implicit self-esteem recommended that John Proudfoot receive a 1-year suspension, whereas participants with relatively high implicit self-esteem recommended he receive only a one-term probation—a nontrivial difference. Individuals with high explicit but low implicit self-esteem may thus be more likely than individuals with high explicit and implicit self-esteem to use ethnic discrimination as a means of defending their positive self-views. When their self-views were challenged by negative performance feedback, participants with high explicit self-esteem and low implicit self-esteem recommended a significantly more severe punishment for John Proudfoot than did participants with high implicit self-esteem. In contrast, implicit self-esteem was unrelated to the severity of punishment recommended for the same offense committed by John Pride, a White offender.

STUDY 2

Although encouraging, the results of Study 1 leave open the possibility that low implicit self-esteem is related to increased discrimination for everyone, not just for individuals with high explicit self-esteem. In other words, anyone with low implicit self-esteem may be prone to discrimination, not just individuals with high explicit but low implicit self-esteem. In Study 2 we thus

conducted an exact replication of Study 1 but drew participants from the full range of explicit self-esteem in order to test whether the negative relation between discrimination and implicit self-esteem observed in Study 1 holds only for individuals with high explicit self-esteem.

Method

PARTICIPANTS

One hundred and thirteen White students enrolled in introductory psychology at the University of Waterloo participated in exchange for partial course credit. Eight participants' data were excluded from analyses because their error rates on the IAT exceeded 20% (Greenwald et al., 1998). This left a total of 105 participants (34 male, 71 female).

MATERIALS AND PROCEDURE

The materials and procedure of Study 2 were identical to those of Study 1. The only difference between the two studies was that in Study 1 only participants who scored in the top third of the RSES distribution participated, whereas in Study 2, participants were not preselected on the basis of their RSES scores.

Results and Discussion

RSES and IAT scores were not related to one another ($r = -.07$, *ns*). As in Study 1, the three items assessing the recommended severity of punishment showed reasonable internal consistency (standardized item $\alpha = .68$). We thus converted each item to z-score form and aggregated the items into a single measure of severity of punishment.

In this study we expected to observe a negative relation between implicit self-esteem and the severity of punishment recommended for John Proudfoot (but not John Pride) among participants with high explicit self-esteem, replicating the findings of Study 1. We also anticipated that this interaction would be attenuated or absent for participants with relatively low levels of explicit self-esteem. This pattern of results would be indicated by a significant three-way interaction between implicit self-esteem, explicit self-esteem, and target ethnicity, such that the interaction between implicit self-esteem and target ethnicity is significant for participants with high explicit self-esteem but not for those with relatively low explicit self-esteem. We thus regressed RSES scores, IAT scores, target ethnicity (dummy coded; White = 0, Native = 1), and all higher order interactions among these variables onto severity of punishment after first centering IAT and RSES scores. Contrary to predictions, this analysis did not reveal a significant three-way interaction, $t(97) < 1$.

Nevertheless, further inspection of the data suggested they might yet support our predictions. Specifically, the

expected two-way interaction between target ethnicity and implicit self-esteem appeared to hold for participants with high explicit self-esteem but not for participants with moderate or low explicit self-esteem. The form of the interaction, however, appeared to change in a somewhat complex, curvilinear manner, as described below. We thus incorporated a curvilinear component of explicit self-esteem into our analysis. To do so, we created a vector representing the quadratic component of explicit self-esteem by squaring participants' explicit self-esteem scores. We then constructed the relevant higher-order interaction terms using this quadratic vector. This analysis revealed a significant three-way interaction between target ethnicity, implicit self-esteem, and the quadratic component of explicit self-esteem, $t(93) = -3.40$, $p = .001$ ($\beta = -.50$). Note that in all analyses we included all lower-order linear and curvilinear terms when testing and interpreting higher-order effects (Cohen, Cohen, West, & Aiken, 2003).

To explore the nature of the three-way interaction further, we tested the two-way interaction between target ethnicity and implicit self-esteem separately for participants with high (+1 *SD*), moderate (mean level), and low (-1 *SD*) levels of explicit self-esteem. As predicted, this two-way interaction was significant for participants with high explicit self-esteem, $t(93) = -2.11$, $p = .038$ ($\beta = -.55$), replicating our Study 1 finding. As in Study 1, for participants with high explicit self-esteem, IAT scores were not significantly related to the severity of punishment recommended for John Pride, $t(93) = 1.10$, $p = .274$ ($\beta = .25$; $y'_{\text{low}} = .15$, $y'_{\text{high}} = .23$). In contrast, among these participants, there was a marginally significant negative relation between implicit self-esteem and severity of punishment for participants who read about John Proudfoot, $t(93) = -1.82$, $p = .07$ ($\beta = -.50$; $y'_{\text{low}} = .69$, $y'_{\text{high}} = -.66$). Probing the interaction from a different perspective, high explicit self-esteem participants with relatively low implicit self-esteem recommended more severe punishments for John Proudfoot than John Pride, although this difference was not significant, $t(93) = 1.23$, *ns*. In contrast, participants with relatively high implicit self-esteem tended to recommend a less severe punishment for John Proudfoot than for John Pride, $t(93) = -1.82$, $p = .07$.

In contrast, the two-way interaction between target ethnicity and implicit self-esteem was not significant among participants with moderate levels of explicit self-esteem, $t(93) = 1.50$, $p = .14$ ($\beta = .29$). In addition, the trend toward an interaction for these participants assumed a rather different form than for individuals with high explicit self-esteem. In this case, there was a marginally significant negative relation between implicit self-esteem and severity of punishment for participants who

read about John Pride, $t(93) = -1.77, p = .08$ ($\beta = -.34$; $y'_{\text{low}} = .12, y'_{\text{high}} = -.39$), and no relation for participants who read about John Proudfoot, $t(93) = .288, p = .77$ ($\beta = .05$; $y'_{\text{low}} = .10, y'_{\text{high}} = .10$). To the extent that implicit self-esteem was related to discrimination, it resulted from higher levels of discrimination among individuals with high implicit self-esteem. Such discrimination, however, reflects in-group favoritism (less severe punishments recommended for a White offender) rather than out-group derogation. Thus, among individuals with moderate levels of explicit self-esteem, implicit self-esteem may be associated with greater ingroup bias, as was originally suggested by Greenwald and Banaji (1995; Farnham et al., 1999).

Finally, among participants with low explicit self-esteem, there was no evidence of a two-way interaction between target ethnicity and implicit self-esteem, $t(93) = -.239, p = .81$ ($\beta = -.05$). Thus, among these participants, implicit self-esteem was not related to discrimination. Although participants with relatively low implicit self-esteem tended to recommend a more severe punishment for John Proudfoot, $t(93) = -1.86, p = .07$ ($\beta = -.43$; $y'_{\text{low}} = .74, y'_{\text{high}} = -.60$), they also tended to recommend a more severe punishment for John Pride, $t(93) = -1.75, p = .08$ ($\beta = -.36$; $y'_{\text{low}} = .32, y'_{\text{high}} = -.76$). Thus, low implicit self-esteem was not associated with greater discrimination among individuals with low explicit self-esteem because people with relatively low implicit self-esteem were more punitive toward both an out-group member and an in-group member. This finding is consistent with the findings of Crocker et al. (1987) for individuals with low explicit self-esteem.

The curvilinear nature of the three-way interaction between explicit self-esteem, implicit self-esteem, and target ethnicity can be seen by observing the regression coefficients for the two-way interaction between implicit self-esteem and target ethnicity across levels of explicit self-esteem. For individuals with high explicit self-esteem, the interaction is negative ($\beta = -.55$); for individuals with moderate explicit self-esteem, the interaction is positive ($\beta = .29$); and for individuals with low explicit self-esteem, the interaction, although negligible, is negative ($\beta = -.05$). Thus, the sign of the two-way interaction changes in a curvilinear manner across levels of explicit self-esteem. Although we did not anticipate the relatively complex form of this three-way interaction, it clearly demonstrates that the negative relation between implicit self-esteem and discrimination found in Study 1 holds only for participants with high explicit self-esteem.

AGGREGATED RESULTS FROM STUDIES 1 AND 2

The results of Study 2 for participants with high explicit self-esteem were thus broadly consistent with

those of Study 1, although some inconsistencies in levels of statistical significance emerged for key contrasts. Because Study 2 was an exact replication of Study 1, we thus aggregated the data of participants in Study 1 with those of participants in the top third of the RSES distribution in Study 2 in order to conduct the most powerful tests available to us of our predictions for individuals with high explicit self-esteem. In this aggregated data set, the two-way interaction between IAT scores and target ethnicity was, not surprisingly, significant, $t(63) = -2.56, p = .01$ ($\beta = -.49$).² Moreover, as can be seen in Figure 1, there was a significant negative relationship between implicit self-esteem and severity of punishment for participants who read about John Proudfoot, $t(63) = -3.14, p = .003$ ($\beta = -.37$), but no relationship between implicit self-esteem and severity of punishment for participants who read about John Pride, $t(63) < 1$ ($\beta = .10$). In addition, testing the difference between the punishments recommended for John Proudfoot and John Pride revealed that participants with low implicit self-esteem (-1 SD) recommended a more severe punishment for John Proudfoot than John Pride, $t(63) = 2.43, p = .018$. There was, however, no significant difference in these ratings for participants with high implicit self-esteem ($+1$ SD), $t(63) = -1.29, p = .20$, although the difference at this point was in the direction of recommending a less severe punishment for John Proudfoot than John Pride.

GENERAL DISCUSSION

Across two studies, we thus found strong support for our prediction that individuals with high explicit but low implicit self-esteem are particularly likely to engage in ethnic discrimination as a means of protecting threatened self-views. Individuals with high explicit self-esteem who had their positive self-views challenged by negative performance feedback recommended a more severe punishment for a Native student who started a fistfight to the extent that they had relatively low implicit self-esteem. This was not the case when a White student committed the same offence. Moreover, when the data for participants with high explicit self-esteem from both studies were aggregated, the results showed clearly that participants with high explicit but low implicit self-esteem recommended a significantly more severe punishment for a Native student than for a White student. If anything, participants with high explicit and high implicit self-esteem showed the opposite tendency. Thus, among individuals with high explicit self-esteem, only those with relatively low implicit self-esteem discriminated ethnically when their self-views were threatened. In addition, Study 2 demonstrated that this pattern of results depends on levels of explicit self-esteem.

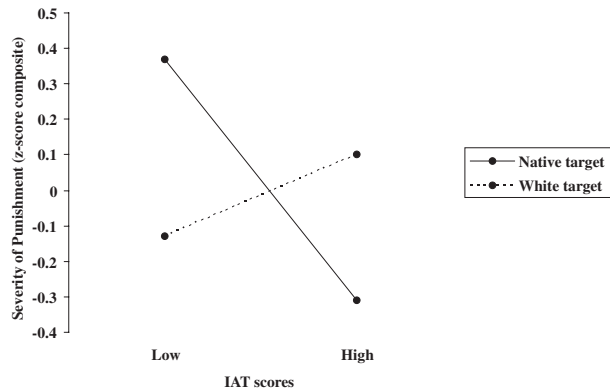


Figure 1 Severity of recommended punishment as a function of implicit self-esteem and target ethnicity for participants with high explicit self-esteem (Studies 1 and 2).

NOTE: IAT = Implicit Associations Test.

Implicit self-esteem was negatively related to discrimination only for individuals with high explicit self-esteem, not for those with moderate or low levels of explicit self-esteem. Thus, it is not the case that low implicit self-esteem is always associated with a greater tendency to discriminate. Taken together, these results thus reinforce the conclusion that individuals with high explicit but low implicit self-esteem are relatively defensive, whereas individuals with high explicit and high implicit self-esteem are more secure (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003).

Nonetheless, it is remarkable that individuals with high explicit and high implicit self-esteem showed no evidence at all of ethnic discrimination. In fact, in Study 2, with participants from the full range of explicit self-esteem, only individuals with high explicit and high implicit self-esteem tended to recommend a less severe punishment for a Native than for a White offender. Fein and Spencer (1997) noted that

stereotyping or derogation of a member of a stereotyped group can provide . . . situational opportunities to restore a threatened self-image. Because it is likely that people often will encounter others in situations where it is personally and socially acceptable to evaluate them negatively, stereotyping and prejudice may be common reactions to self-image threat. (p. 40)

In the present study, the situation was engineered such that it would be socially acceptable to derogate an out-group member—the student had, after all, behaved in a clearly violent manner and ought to be punished. In addition, participants with high explicit self-esteem arguably faced the greatest potential threat to their self-views, because they were initially so high. Yet, participants with

high explicit and high implicit self-esteem eschewed the opportunity to enhance themselves by discriminating against an ethnic minority member. This might suggest that such individuals did not view ethnic discrimination as a personally acceptable means of defending their self-views.

In fact, discriminating ethnically can challenge some individuals' self-images of being fair and equitable (see Abrams & Hogg, 1988). People with high explicit and implicit self-esteem may be particularly unlikely to act contrary to their personal morals and values in order to maintain their self-views, because behaving contrary to their morals could further damage their self-esteem. Such individuals do report that their self-worth depends on behaving in line with their own values (Jordan, Spencer, & Zanna, 2003). Moreover, individuals with high explicit and implicit self-esteem may be less likely than individuals with high explicit but low implicit self-esteem to place primacy on the goal of maintaining positive self-views and so may be less willing to compromise other important goals—such as the goal of behaving equitably or of maintaining positive social relations—in order to feel good about themselves (Crocker & Park, 2003; Epstein & Morling, 1995). It is also possible, however, that individuals with high explicit and implicit self-esteem did not discriminate because they were not overly threatened by the negative feedback or because they had more self-affirmational resources to draw on in order to neutralize the threat without resorting to discrimination (Fein & Spencer, 1997).

In any event, the present findings suggest that high self-esteem can assume qualitatively different forms that are related to prejudiced and discriminatory behavior. Our preferred interpretation of these findings is that the threat posed by negative performance feedback brought self-doubts and insecurities to mind for individuals with high explicit but low implicit self-esteem, which then motivated discrimination as a defensive response. Most conceptualizations of implicit self-esteem view it as existing entirely outside awareness (e.g., Farnham et al., 1999; Pelham & Hetts, 2001). We believe, however, that it is *preconscious*, sometimes entering consciousness. When individuals with high explicit but low implicit self-esteem have their positive self-views challenged or threatened, we believe that their low implicit self-esteem enters awareness. That is, they experience relatively negative self-feelings that contradict their normally positive self-views, particularly compared with individuals with high explicit and high implicit self-esteem. Such experiences, moreover, may motivate people with high explicit but low implicit self-esteem to confirm their explicitly positive self-views and deny their implicitly negative self-views through a variety of self-esteem maintenance techniques

(for a more detailed discussion, see Jordan, Spencer, & Zanna, 2003; Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). In the present studies, we believe they discriminated against a Native student offender in order to protect their challenged self-views. The present data, however, do not let us determine whether implicit self-esteem is unconscious, preconscious, or fully conscious in nature.

In addition, the present studies leave another intriguing question unanswered. What role does self-threat play in these effects? When designing the present studies, we reasoned that the effect would emerge most clearly when participants' self-views are threatened because Fein and Spencer (1997), using similar methods, found greater negative stereotyping only in response to ego threat. Without manipulating threat directly in the present context, however, it is not clear whether it is a necessary precursor to discrimination. Indeed, it is possible that defensive individuals are more likely than secure individuals to derogate out-group members even in the absence of immediate threat. This might be the case if they have learned through past discriminatory acts that such behavior upholds their positive self-views, and thus they habitually assert their superiority over ethnically different others in situations where it is perceived to be acceptable to do so. In keeping with this possibility, it has been found that narcissists, even in the absence of threat, appear more prejudiced than do nonnarcissists (Brown, 2000), and individuals with high explicit but low implicit self-esteem tend to be relatively narcissistic (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). Conversely, comparing oneself to an out-group member may be inherently threatening to defensive individuals, and so they may habitually discriminate when the situation engenders such comparison. This latter possibility is consistent with the finding that individuals with high explicit but low implicit self-esteem engage in more in-group bias in the minimal group paradigm than do individuals who are high in both types of self-esteem (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). On the other hand, among individuals with high explicit self-esteem, those with relatively low implicit self-esteem do not appear to be higher in either right wing authoritarianism or social dominance orientation (Jordan, Zanna, & Spencer, 2003), the two strongest predictors of generalized prejudice (Altemeyer, 1998), suggesting that such individuals might only discriminate more in responses to ego threat.

Overall, then, the present results contribute to a growing body of evidence that high explicit self-esteem is heterogeneous. Individuals with high explicit but low implicit self-esteem appear to be relatively defensive, whereas those with high explicit and high implicit self-

esteem appear to be relatively secure. In our past research, we found that the former individuals were more narcissistic, rationalized their decisions more, and favored in-group members more in the minimal group paradigm (Jordan, Spencer, Zanna, Hoshino-Browne, et al., 2003). The present studies extend these findings by suggesting that people with high explicit but low implicit self-esteem may also be particularly likely to discriminate on the basis of ethnicity in order to defend their own positive self-views. In this sense, individuals with high explicit but low implicit self-esteem do not seem to be much like Willie Nelson. They do not seem to be particularly gracious, and they may, at least sometimes, "take from others to be strong."

NOTES

1. y'_{high} and y'_{low} represent predicted scores on the severity of punishment measure predicted from the overall regression equation at values of one standard deviation above and below the mean of implicit self-esteem, respectively (Cohen, Cohen, West, & Aiken, 2003).

2. We also ran an analysis including the source of data as a factor (whether the data were from Study 1 or 2). This analysis showed that the observed two-way interaction between Implicit Associations Test (IAT) scores and target race was not moderated by the source of the data, $t(59) < 1$.

REFERENCES

- Aberson, C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review, 4*, 157-173.
- Abrams, D., & Hogg, M. A. (1988). Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology, 18*, 317-334.
- Altemeyer, B. (1998). The other "authoritarian personality." in M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 47-92). San Diego, CA: Academic Press.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1-44.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631-643.
- Brown, R. P. (2000, February). *Of pride and prejudice: Narcissism, self-esteem, and the prejudice-insecurity hypothesis*. Poster presented at the second annual meeting of the Society for Personality and Social Psychology, San Antonio, TX.
- Bushman, B. J., & Baumeister, R. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*, 219-229.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Coopersmith, S. (1959). A method for determining types of self-esteem. *Journal of Abnormal and Social Psychology, 59*, 87-94.
- Crocker, J., & Park, L. E. (2003). Seeking self-esteem: Construction, maintenance, and protection of self-worth. In M. Leary and J. Tangney (Eds.), *Handbook of self and identity* (pp. 291-313). New York: Guilford.
- Crocker, J., Thompson, L., McGraw, K., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others:

- Effects of self-esteem and threat. *Journal of Personality and Social Psychology*, 52, 907-916.
- Deci, E. L., & Ryan, R. M. (1995). Human autonomy: The basis for true self-esteem. In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 31-49). New York: Plenum.
- Emler, N. (2001). *Self-esteem: The costs and causes of low self-worth*. York, UK: York Publishing Services for the Joseph Rowntree Foundation.
- Epstein, S., & Morling, B. (1995). Is the self motivated to do more than enhance and/or verify itself? In M. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 9-29). New York: Plenum.
- Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. Mackie & D. L. Hamilton (Eds.), *Affect, cognition and stereotyping: Interactive processes in group perception* (pp. 137-166). New York: Academic Press.
- Esses, V. M., & Zanna, M. P. (1995). Mood and the expression of ethnic stereotypes. *Journal of Personality and Social Psychology*, 69, 1052-1068.
- Farnham, S. D., Greenwald, A. G., & Banaji, M. R. (1999). Implicit self-esteem. In D. Abrams & M. A. Hogg (Eds.), *Social identity and social cognition* (pp. 230-248). Oxford, UK: Basil Blackwell.
- 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.
- 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. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Haddock, G., & Zanna, M. P. (1994). The (limited) role of trait-laden stereotypes in predicting attitudes toward Native peoples. *British Journal of Social Psychology*, 33, 83-106.
- Harder, D. W. (1984). Character style of the defensively high self-esteem man. *Journal of Clinical Psychology*, 40, 26-35.
- Heatherton, T. F., & Vohs, K. D. (2000). Interpersonal evaluations following threats to the self: Role of self-esteem. *Journal of Personality and Social Psychology*, 78, 725-736.
- Hogg, M. A., & Abrams, D. (1990). Social motivation, self-esteem, and social identity. In D. Abrams & M. A. Hogg (Eds.), *Social identity theory: Constructive and critical advances* (pp. 28-47). New York: Harvester Wheatsheaf.
- Horney, K. (1937). *The neurotic personality of our time*. New York: Norton.
- Jordan, C. H., Spencer, S. J., & Zanna, M. P. (2003). "I love me . . . I love me not": Implicit self-esteem, explicit self-esteem, and defensiveness. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *The Ontario Symposium: Vol. 9. Motivated social perception* (pp. 117-145). Mahwah, NJ: Lawrence Erlbaum.
- Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, 85, 969-978.
- Jordan, C. H., Zanna, M. P., & Spencer, S. J. (2003). *Implicit self-esteem, explicit self-esteem and their relation to right wing authoritarianism and social dominance orientation*. Unpublished raw data, University of Waterloo, Ontario, Canada.
- Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin*, 30, 22-34.
- Keenan, D. (2003). Carla Bozulich. *The Wire*, 236, 14.
- Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry*, 4, 1-26.
- Kernis, M. H., & Paradise, A. W. (2002). Distinguishing between secure and fragile forms of high self-esteem. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 339-360). Rochester, NY: University of Rochester Press.
- Pelham, B. W., & Hetts, J. J. (2001). A case for the nonconscious self-concept. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 105-123). Mahwah, NJ: Lawrence Erlbaum.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processing in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32, 880-892.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108-131.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90, 245-271.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.

Received December 1, 2003

Revision accepted September 1, 2004