

Research Report

Increasing Implicit Self-Esteem Through Classical Conditioning

Jodene R. Baccus, Mark W. Baldwin, and Dominic J. Packer

McGill University, Montreal, Quebec, Canada

ABSTRACT—*Implicit self-esteem is the automatic, nonconscious aspect of self-esteem. This study demonstrated that implicit self-esteem can be increased using a computer game that repeatedly pairs self-relevant information with smiling faces. These findings, which are consistent with principles of classical conditioning, establish the associative and interpersonal nature of implicit self-esteem and demonstrate the potential benefit of applying basic learning principles in this domain.*

Low self-esteem has been implicated in a host of consequential social phenomena, including drug abuse, hostility, and relationship dysfunction (Harter, 1993; Leary, Schreindorfer, & Haput, 1995). A person's level of self-esteem is typically viewed as the sum of his or her conscious self-evaluative thoughts and feelings. Therefore, it is most often assessed explicitly through self-report scales containing items such as "At times I think I am no good at all" (Rosenberg, 1965), and efforts to enhance explicit self-esteem focus on altering self-critical thoughts (Brewin, 1989).

Recent social cognitive research into the multifaceted nature of self-esteem, however, has highlighted an automatic component dubbed *implicit self-esteem*, conceptualized as a self-evaluation that occurs unintentionally and often outside of awareness (Farnham, Greenwald, & Banaji, 1999). Implicit self-esteem can be thought of as an automatic attitude toward the self, which influences subsequent evaluations of the self and of self-relevant objects in the environment. Because it is not dependent on explicit awareness and so in principle is unsuited to direct measurement by self-report, researchers have developed measures to assess implicit self-esteem through indirect means, for example, by examining automatic associations between "self" and "good," or by assessing evaluative responses to self-relevant objects (Greenwald & Farnham, 2000; Jones, Pelham, & Mirenberg, 2002; Kitayama & Karasawa, 1997; Koole, Dijksterhuis, & van Knippenberg, 2001). Implicit and explicit self-esteem are typically weakly correlated, at best, confirming that they reflect largely independent processes (Bosson, Swann, & Pennebaker, 2000; Farnham et al., 1999; Greenwald & Farnham, 2000). Notably, research has shown that implicit self-esteem is better than explicit self-esteem in

predicting observer-rated nonverbal anxiety and negative mood in response to threatening feedback, as well as persistence at tasks and socially undesirable behavior in response to failure (Bosson et al., 2000; Greenwald & Farnham, 2000; Jordan, Spencer, & Zanna, 2002; Pelham & Hetts, 1999; Spalding & Hardin, 1999).

Presently, the cognitive foundations of implicit self-esteem remain largely unexplored. We investigated the role played by fundamental processes of associative learning. According to general theories of attitude formation, whereas explicit attitudes develop via persuasion, rational argument, and other verbal means, implicit attitudes develop primarily via the repeated pairings of potential attitude objects with positive and negative goal-relevant stimuli (Karpinski & Hilton, 2001; Olson & Fazio, 2001; Staats & Staats, 1958; Walther, 2002). Drawing on these theories, we posited that implicit self-esteem arises from associative links between the self-concept and positive versus negative social information. To test this conceptualization, we sought to modify the underlying associations.

We drew on elementary Pavlovian conditioning, wherein two stimuli are repeatedly paired until the presence of one evokes the expectation of the other (Dickinson, 1989; Pavlov, 1927), an approach that has been used to modify attitudes toward objects in the environment (De Houwer, Thomas, & Baeyens, 2001; Olson & Fazio, 2001; Walther, 2002). Whereas previous research had focused on attitudes toward external objects, we paired positive and negative stimuli with an internal construct—the self—and examined the effects of this pairing on implicit self-esteem (see also Dijksterhuis, 2004, for related research on self-attitudes). Previous research had shown that self-esteem is largely interpersonally based, with positive thoughts and feelings about the self arising from the sense of being securely accepted and positively regarded by others (Baldwin & Baccus, 2003; Baldwin & Sinclair, 1996; Dandeneau & Baldwin, in press; Leary, Tambor, Terdal, & Downs, 1995). Thus, we devised a computer game to repeatedly pair self-relevant information with positive social feedback (photographs of smiling faces) to create an automatic anticipation of secure social acceptance and consequently enhanced self-evaluations on implicit measures. We also tested the impact of the manipulation on a phenomenon of social relevance: aggression. Because feelings of low self-esteem based in social insecurity have been linked to aggressive behavior (Kirkpatrick, Waugh, Valencia, & Webster, 2002), we predicted individuals with low self-esteem would show less evidence of self-reported aggressive thoughts and feelings if they underwent the self-acceptance conditioning procedure than if they were in a control condition.

Address correspondence to Jodene Baccus, Department of Psychology, McGill University, 1205 Docteur Penfield Ave., Montreal, QC H3A 1B1, Canada; e-mail: jodene.baccus@mail.mcgill.ca.

METHOD

Participants

Participants were 139 undergraduate volunteers (mean age = 19.1 years) from McGill University and Dawson College in Montreal, Canada. We discarded data from 20 participants who did not fully complete one of the implicit measures and 1 participant whose score on the other implicit measure was greater than 3 standard deviations above the mean, leaving us with a final sample of 118 (36 men, 82 women).

Materials and Procedure

Participants first completed the Rosenberg Self-Esteem Scale (Rosenberg, 1965) as a premanipulation measure of explicit self-esteem and the Name Letter measure (described later in Method) as a premanipulation measure of implicit self-esteem. They were then seated at a computer and randomly assigned to either the experimental or the control version of the conditioning task. All participants entered into the computer the answers to six questions about themselves (e.g., first name, date of birth) and were then instructed that a word would appear randomly in one of the quadrants on the computer screen and their task was to click on the word as quickly as possible (see Fig. 1). They were also told that doing so would cause an image to be displayed briefly (for 400 ms) in that quadrant. This procedure was repeated for 240 trials. The words presented were chosen from the participant's answers to the six questions at the start of the session (self-relevant words), as well as from a preprogrammed list of words fitting the same categories (non-self-relevant words). In the control condition, a random selection of smiling, frowning, and neutral photographs of men and women followed both self-relevant (80 trials) and non-self-relevant words. In the experimental condition, self-relevant words were always paired with an image of a smiling face. Participants in the two conditions received identical numbers of each kind of expression. Following the computer game, participants completed the Implicit Association Test (IAT; described later in Method) and the Name Letter measure.

After the conditioning task, participants also completed the State Self-Esteem Scale (Heatherton & Polivy, 1991), the Profile of Mood States (McNair, Lorr, & Droppleman, 1971), and an aggression measure (adapted from Taylor, 1967). For the latter measure, participants were asked to imagine being in three situations in which they were competing with another student. Some of these scenarios involved being insulted or rejected by the other person. In each scenario, they won the competition and were allowed to "punish" the other student with a blast of noise. They were asked to rate how loud and how long they felt like setting, and would set, the noise blast. We calculated a total aggression score ($\alpha = .90$) by collapsing across the three scenarios.

Implicit Self-Esteem Measures

Name Letter Measure

In the Name Letter measure, participants rate their liking for each letter of the alphabet. High implicit self-esteem is indexed by the extent to which a person prefers his or her initials to other letters of the alphabet (Jones et al., 2002; Kitayama & Karasawa, 1997). To control for response styles involving the tendency to use high or low numbers on the scale, we first ipsatized ratings by subtracting from each participant's ratings of his or her initials the mean liking score that participant gave to the remaining letters of the alphabet. To then

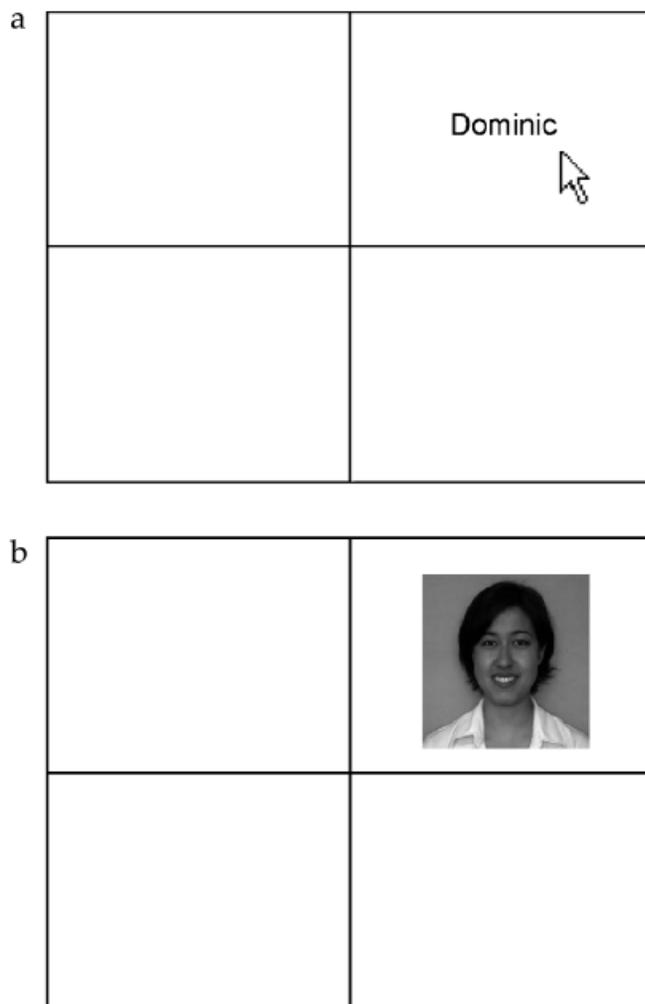


Fig. 1. The conditioning task. When a word appeared in one quadrant of the screen (a), participants clicked on the word with the computer mouse as quickly as possible. After the click, a photograph of a smiling, frowning, or neutral face (b) was shown for 400 ms.

control for a potential confound that certain frequently used letters might generally be rated higher than other less frequent letters (Jones et al., 2002), we subtracted from the rating of each of these letters the mean ipsatized score for all other participants who did not share that initial. Each participant's score was the mean of the adjusted ratings for his or her two initials.

Implicit Association Test

The IAT (Greenwald & Farnham, 2000) is a reaction time task that requires participants to sort words into categories. In one block of trials, the correct response for self-related words (e.g., *me*) and the correct response for pleasant words (e.g., *rainbow*) involved pressing the same computer key; in another block, self-related words were assigned the same key as unpleasant words (e.g., *vomit*). Faster reaction times are theorized to reflect stronger associations between the types of words sharing a response key. High implicit self-esteem, therefore, was indexed by the amount of time it took participants to respond to the target words when self-related and positive words shared the same key, relative to the amount of time it took to respond

to the same target words when self-related and negative words shared a key. Recent research has suggested that responses to “self” targets and “other” targets might not reflect the same underlying construct (Karpinski & Kiefer, 2002), however, so we examined self-targets as the most direct indicator of implicit self-esteem.

Implicit Self-Esteem Composite Score

As expected (Bosson et al., 2000), the postmanipulation Name Letter and IAT scores were uncorrelated in our sample ($r = .061$, n.s.). Because it has been argued that it is best to triangulate on the construct of implicit self-esteem by combining multiple measurement approaches (see, e.g., Cunningham, Preacher, & Banaji, 2001, for confirmatory factor analyses demonstrating that multiple measures of implicit attitudes converge on a single construct), we computed a composite value of the two measures for each participant by taking the mean of the participant’s z scores.

RESULTS

Preliminary Analyses

As in earlier research (e.g., Bosson et al., 2000), the premanipulation measures of explicit self-esteem and implicit self-esteem were uncorrelated ($r = -.020$, n.s.), confirming that they represent independent aspects of self-esteem.

Primary Analyses: Composite Scores

We conducted multiple regression analyses to examine the effects of the conditioning task, either alone or in interaction with premanipulation implicit or explicit self-esteem, on the composite score for postmanipulation implicit self-esteem. Our primary hypothesis was supported in that the computer game was successful at enhancing implicit self-esteem. Participants completing the experimental version of the conditioning task exhibited significantly higher implicit self-esteem than those completing the control version, $t(110) = 2.80$, $p = .006$ (see Fig. 2). There were no effects involving gender.

There was also a significant three-way interaction between condition, premanipulation implicit self-esteem, and premanipulation explicit self-esteem, $t(110) = 2.81$, $p = .006$. Analyses of the simple slopes (following Aiken & West, 1991) revealed that participants who began the study with congruent implicit and explicit self-esteem (low in both implicit and explicit self-esteem or high in both) showed the most pronounced increases in self-esteem following the conditioning task, $t(110) = 2.99$, $p = .003$, and $t(110) = 2.48$, $p = .015$, for those with low implicit and low explicit self-esteem and those with high implicit and high explicit self-esteem, respectively. Parallel analyses based on groups created with median splits produced similar effects; for ease of interpretation, these means are displayed in Figure 2.¹

¹We conducted separate analyses on the two postmanipulation measures of implicit self-esteem. Analyses on the Name Letter measure produced a two-way interaction between premanipulation implicit self-esteem and condition, $t(110) = -2.87$, $p = .005$; the beneficial effect of the manipulation was most pronounced for participants who began the study with low implicit self-esteem, $t(110) = 3.22$, $p = .002$. Analyses of the IAT showed only the overall main effect of the manipulation; participants in the experimental condition had significantly higher implicit self-esteem scores than participants in the control condition, $t(109) = 2.01$, $p = .047$. Even when examined separately, then, both measures showed an impact of the manipulation on implicit self-esteem.

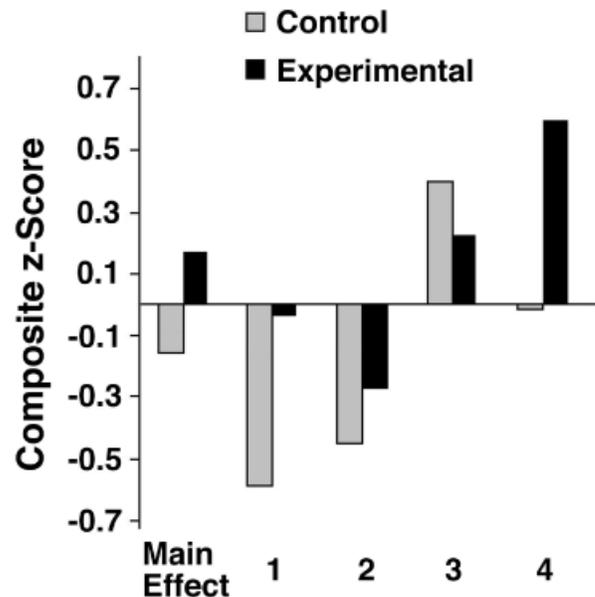


Fig. 2. Postmanipulation implicit self-esteem (mean of z scores on the Name Letter measure and Implicit Association Test) in the control and experimental conditions. Note that because the composite was calculated from z scores, the grand mean is zero. The graph displays the overall main effect of condition and the condition effect in the four groups defined by median splits of premanipulation implicit and explicit self-esteem: 1 = low implicit and low explicit self-esteem; 2 = low implicit and high explicit self-esteem; 3 = high implicit and low explicit self-esteem; 4 = high implicit and high explicit self-esteem.

Postmanipulation implicit self-esteem was not correlated with postmanipulation explicit self-esteem ($r = -.035$, n.s.). Furthermore, there were no statistically significant effects on postmanipulation explicit state self-esteem involving condition (all $ps > .118$). Implicit self-esteem was not correlated with mood overall ($r = .001$, n.s.). However, regression analysis of mood revealed a two-way interaction between condition and premanipulation implicit self-esteem, $t(106) = -2.34$, $p = .021$, such that participants who began the study high in implicit self-esteem showed a more positive mood in the control condition than in the experimental condition. Possibly random, unpredictable exposure to smiling faces boosted mood among individuals already high in self-esteem by activating a generally positive orientation toward the environment. This pattern is different from the effect of condition on postmanipulation implicit self-esteem, and the postmanipulation increases in implicit self-esteem remained significant when explicit self-esteem and mood were statistically controlled.

Secondary Analyses: Aggression

To explore the effects of the conditioning task on social behavior, we performed regression analyses on the measure of aggression. The only effect involving condition was a two-way interaction between premanipulation explicit self-esteem and condition, $t(99) = 2.93$, $p = .004$. Tests of the simple slopes showed that participants with low premanipulation explicit self-esteem reported significantly lower aggressive thoughts and feelings after the manipulation if they were in the experimental group than if they were in the control group, $t(99) = -2.53$, $p = .013$.

DISCUSSION

Participants who performed a conditioning task in which they were repeatedly exposed to pairings of self-relevant information with smiling faces showed enhanced implicit self-esteem compared with control subjects. Thus, a simple conditioning paradigm, originally developed by learning theorists to study animal responses to expectations of food or shock, was effective at modifying people's unconscious responses to themselves. This finding fits well with theories that situate the roots of self-acceptance in the anticipation of positive, warm feedback from other people. It also provides evidence that the feelings of low self-esteem harbored by some individuals are not set in stone in childhood, but might be raised at a later time via an intervention applying basic learning principles.

We found that the conditioning manipulation tended to be most effective among participants with congruent implicit and explicit self-esteem. Although this moderation of the conditioning effect was not predicted and requires replication, it is consistent with other findings in the literature suggesting that individuals with incongruent implicit and explicit self-esteem are more likely to exhibit troublesome self-esteem dynamics such as narcissism and prejudice than are individuals with congruent implicit and explicit self-esteem (e.g., Jordan et al., 2002). We speculate that certain forms of incongruent or defensive self-esteem might interfere with people's ability to benefit from positive social feedback.

The impact of the conditioning extended beyond automatic self-evaluative reactions to feelings of aggressiveness. Participants who began the study with low explicit self-esteem and subsequently completed the conditioning task showed significantly lower levels of aggressiveness compared with their counterparts in the control condition. These initial findings require extension to actual social behavior, but we find them promising. Some computer and video games have been subject to criticism based on the possibility that they might reinforce antisocial attitudes and behaviors among game players. The current findings suggest that a game involving a specific pattern of repetitive exposures to positive social feedback might, conversely, lead to favorable outcomes.

Overall, the current research provides insight into the cognitive foundations of implicit self-esteem, suggesting that implicit self-esteem reactions are rooted in fundamental associations between the self-representation and expectations of positive versus negative social feedback. Recently, self-theorists have proposed that the self-esteem system functions automatically to assess the likelihood of acceptance versus exclusion by other people. Our findings demonstrate that the system also stores this information in the form of positive or negative associations to the self, and these associations can be modified via an acceptance-conditioning procedure.

Acknowledgments—This research was supported by a fellowship to the first author and a grant to the second author from the Social Sciences and Humanities Research Council of Canada. We thank John Lydon, Karim Nader, and David Paul for comments, and Christopher Bryan, Eleshia Morrison, Maya Sakellaropoulou, and Genevieve Taylor for their help with data collection.

REFERENCES

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

- Baldwin, M.W., & Baccus, J.R. (2003). An expectancy value approach to self-esteem. In S.J. Spencer, S. Fein, M.P. Zanna, & J.M. Olson (Eds.), *The Ontario Symposium: Vol. 9. Motivated social perception* (pp. 171–194). Mahwah, NJ: Erlbaum.
- Baldwin, M.W., & Sinclair, L. (1996). Self-esteem and “if...then” contingencies of interpersonal acceptance. *Journal of Personality and Social Psychology, 71*, 1130–1141.
- 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.
- Brewin, C. (1989). Cognitive change processes in psychotherapy. *Psychological Review, 96*, 379–394.
- Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163–170.
- Dandaneau, S.D., & Baldwin, M.W. (in press). The inhibition of socially rejecting information among people with high versus low self-esteem: The role of attentional bias and the effects of bias reduction training. *Journal of Social and Clinical Psychology*.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853–869.
- Dickinson, A. (1989). Expectancy theory in animal conditioning. In S.B. Klein & R.R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 279–308). Hillsdale, NJ: Erlbaum.
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345–355.
- Farnham, S.D., Greenwald, A.G., & Banaji, M.R. (1999). Implicit self-esteem. In D. Abrams & M.A. Hogg (Eds.), *Social cognition and social identity* (pp. 230–248). London: Blackwell.
- Greenwald, A.G., & Farnham, S.D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.
- Harter, S. (1993). Causes and consequences of low self-esteem in children and adolescents. In R.F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 87–116). New York: Plenum Press.
- Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*, 895–910.
- Jones, J.T., Pelham, B.W., & Mirenberg, M.C. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology, 38*, 170–177.
- Jordan, C.H., Spencer, S.J., & Zanna, M.P. (2002, February). *Self-esteem and defensiveness: The roles of implicit and explicit self-esteem*. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Savannah, GA.
- Karpinski, A., & Hilton, J.L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*, 774–788.
- Karpinski, A., & Kiefer, A.K. (2002, February). *The Implicit Association Test as a measure of self-esteem: The role of the other*. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Savannah, GA.
- Kirkpatrick, L.A., Waugh, C.E., Valencia, A., & Webster, G.D. (2002). The functional domain specificity of self-esteem and the differential prediction of aggression. *Journal of Personality and Social Psychology, 82*, 756–767.
- Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin, 23*, 736–742.
- Koole, S.L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80*, 669–685.
- Leary, M.R., Schreindorfer, L.S., & Haput, A.L. (1995). The role of low self-esteem in emotional and behavioral problems: Why is low self-esteem dysfunctional? *Journal of Social and Clinical Psychology, 14*, 297–314.

- Leary, M.R., Tambor, E.S., Terdal, S.K., & Downs, D.L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology, 68*, 518–530.
- McNair, D., Lorr, M., & Droppleman, L. (1971). *Profile of Mood States manual*. San Diego, CA: Educational and Industrial Testing Service.
- Olson, M.A., & Fazio, R.H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417.
- Pavlov, I.P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press.
- Pelham, B.W., & Hetts, J.J. (1999). Implicit and explicit personal and social identity: Toward a more complete understanding of the social self. In T. Tyler & R. Kramer (Eds.), *The psychology of the social self* (pp. 115–143). Mahwah, NJ: Erlbaum.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Spalding, L.R., & Hardin, C.D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10*, 535–539.
- Staats, A.W., & Staats, C.K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal Psychology, 57*, 37–40.
- Taylor, S.P. (1967). Aggressive behavior and psychological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35*, 297–310.
- Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82*, 919–934.

(RECEIVED 12/20/02; REVISION ACCEPTED 7/29/03)