Doing the Implicit Association Test (IAT) Reduces Positive Intergroup Interaction Behavior

Jacquie D. Vorauer

University of Manitoba

Word Count: Main Text (3895) + Footnotes (68) + Acknowledgements (37) = 4000

Address Correspondence to:
Jacquie D. Vorauer
Department of Psychology
University of Manitoba
Winnipeg, Manitoba
Canada, R3T 2N2
email: vorauer@cc.umanitoba.ca
phone: 204-474-8250; fax: 204-474-7599

Abstract

It is frequently suggested that increasing awareness of intergroup bias and limited control over biased responses can improve intergroup interaction behavior. Some uses of the Implicit Association Test (IAT) epitomize this approach to improving intergroup relations. However, if doing the IAT enhances caution and inhibition, reduces self-efficacy, or primes categorical thinking, it may instead have negative effects. Two experiments demonstrated that when White individuals completed a race-relevant IAT prior to an intergroup interaction (as compared to when they did not), their interaction partner left the exchange feeling less positively regarded. No such effects were evident when White individuals completed a race-irrelevant IAT (Study 1) or explicit prejudice measure (Study 2) before the exchange, or when their interaction partner was White (Study 1). Mediation analyses (Study 2) suggested that White participants who did the IAT communicated less positive regard because they adopted a cautious approach to the interaction involving limited self-disclosure.

Doing the Implicit Association Test (IAT) Reduces Positive Intergroup Interaction Behavior

The idea that making people aware of their intergroup biases is a good first step toward reducing those biases has been an increasingly prevalent theme in both the academic literature on intergroup relations (e.g., Devine, 1989; Monteith & Mark, 2005) and the popular media (e.g., Gladwell, 2005; Moss, 2009; Science Daily, 2009) and is core component of many anti-racism campaigns. Some uses of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) epitomize this approach to improving intergroup relations. The IAT involves doing a series of word and picture categorization tasks that ultimately indicate whether individuals associate particular concepts with some groups more strongly than with others (e.g., whether they associate negative concepts with African Americans more strongly than with White Americans).

In line with the belief that heightening awareness of bias is beneficial, researchers have developed an educational website ("Project Implicit") in which individuals can take different versions of the test and receive feedback on their level of bias (https://implicit.harvard.edu). The website has proven to be extremely popular: Many psychology textbooks and media outlets encourage people to take the test, and over five million people worldwide have done so (Nosek et al., 2007).

The website frames the test as a means for individuals to "examine their own hidden biases," and indeed the vast majority (estimates are often in the neighborhood of 80% or more) receive feedback that they are biased. As noted on the website, even without the feedback individuals are likely to notice their own intergroup bias as they respond to different components of the test. Individuals who take the test are also likely to become more aware of their limited control over their responses in the intergroup domain: "The site was designed to allow Web

visitors to experience what the authors and many laboratory subjects have experienced: inability to control the manifestations of automatic associations that are elicited by the IAT method" (Greenwald, Nosek, & Banaji, 2003, p. 198). This limited control over responses and associated minimization of socially desirable responding is one reason that the test has been used in hundreds of psychology studies (Greenwald et al., 2003).

Surprisingly, although the malleability and construct and predictive validity of scores on the test have been the focus of considerable debate and analysis (e.g., Arkes & Tetlock, 2004; Blanton & Jaccard, 2008; Dasgupta & Greenwald, 2001; Greenwald, Poehlman, Uhlmann, & Banaji, 2009), how the act of doing the test actually affects intergroup behavior has received little empirical attention. The present study examines the implications of doing a race-based version of the IAT for behavior toward an outgroup member during a subsequent face-to-face intergroup interaction.

Conceivably, as implied in many discussions of the IAT in particular and prejudice reduction in general, enhancing awareness of bias and limited control over responses could lead individuals to better regulate their intergroup behavior and thus have a salutary effect on the signals they communicate to outgroup members. However, it seems more likely that being alerted to potential bias and limited response control – particularly through a direct and concrete personal experience such as that provided by the IAT – has negative implications for individuals' intergroup interaction behavior.

Several qualities of the IAT could contribute to such negative effects. According to an *enhanced-caution* account, alerting individuals to the possibility of bias could feed into a cautious, prevention-oriented approach to intergroup exchanges that involves limited self-disclosure and engagement: Fear of saying or doing the wrong thing may lead to inhibited

behavior that comes across as unfriendly and aloof (see Devine, Evett, & Vasquez-Suson, 1996; Vorauer & Turpie, 2004). According to a *reduced-efficacy* account, a focus on limited control over responses may undermine feelings of efficacy during intergroup interaction, which are important to positive intergroup contact experiences (e.g., Doerr, Plant, Kunstman, & Buck, 2011; Plant & Butz, 2006). Finally, according to an *enhanced-stereotyping* account, the mere task of categorizing individuals into one of two clearly separate groups, especially when one of them is the ingroup, may detract from the kinds of complex, creative, and inclusive thinking that have been shown to reduce stereotype activation and prompt more positive intergroup behavior (e.g., Bigler & Liben, 1992; Dovidio, Gaertner, Validzic, Matoka, Johnson, & Frazier, 1997; Sassenberg & Moskowitz, 2005).

Two studies probed the behavioral effects of completing the IAT. In each case the key condition was one in which, immediately prior to engaging in a face-to-face interaction with a person with an Aboriginal ethnic background (hereafter referred to as Aboriginal), participants with a European ethnic background (hereafter referred to as White) completed a race-relevant IAT that involved categorizing pictures of White and Aboriginal individuals. The implications of the IAT manipulation for how White participants' Aboriginal partners experienced the exchange were of particular interest. Aboriginal partners' metaperceptions regarding how they were viewed by the White participant were the key dependent measure. Each study included different control conditions to rule out different alternative explanations for the results and Study 2 directly tested enhanced-caution, reduced-efficacy, and enhanced-stereotyping accounts for the effects.

The study designs took a conservative approach in that all participants – even those in the control conditions – were informed of the study's focus on intergroup interaction upon their

arrival. Previous research suggests that such information in and of itself can interfere with intergroup interaction behavior (Vorauer & Turpie, 2004). Further, again to create a stringent test, and also for ethical reasons, participants who did the IAT did not receive feedback on their performance. The mere experience of doing the test should be sufficient to alert individuals to the possibility of bias and their limited response control, especially as the vast majority of people do exhibit bias and make mistakes while doing the test. Moreover, most are able to deduce the purpose of the test even when they are given an elaborate cover story designed to disguise it (Frantz, Cuddy, Burnett, Ray, & Hart, 2004).

Study 1

In Study 1, White participants completed a race-relevant IAT (RR-IAT) that involved categorizing pictures of White and Aboriginal individuals, a race-irrelevant IAT (RI-IAT) that involved categorizing pictures of insects and flowers, or did neither of these tasks, immediately prior to engaging in a face-to-face interaction with a White or Aboriginal partner. The main hypothesis was that Aboriginal partners of White individuals who had completed the RR-IAT would feel more negatively evaluated than would Aboriginal partners of White individuals who had not done the RR-IAT. The intragroup and RI-IAT conditions were included to demonstrate that the effects did not reflect general depletion or feelings of failure.

Method

Participants

Participants were 90 same-sex pairs of previously unacquainted Canadian introductory psychology students (72% female) who received partial course credit. Forty-seven pairs included two White students; the remainder included one White and one Aboriginal student. Students were assigned to pairs on the basis of scheduling convenience. All participants had previously

completed a mass testing survey that included demographic questions and the racial ingroup

identification component of Luhtanen and Crocker's (1992) collective self-esteem scale. This

measure was administered in light of previous research indicating that ingroup identification is

generally associated both with perceived negative evaluations from outgroup members and with

negative evaluations of outgroup members (e.g., Branscombe, Schmitt, & Harvey, 1999; Major,

Quinton, & Schmader, 2003; Kaiser & Wilkins, 2010). The White member of each White-

Aboriginal pair and a randomly selected member of each White-White pair was randomly

assigned to the RR-IAT, RI-IAT, or no-task control condition.

Procedure

Pair members arrived for a study of "social perception in first meeting situations." They

were assigned to wait for the White female experimenter in different locations and were kept

separate at all times except for the discussion and debriefing. As a cover story, the experimenter

told participants that some pairs would interact face-to-face and others would only exchange

written information. In reality all pairs had a face-to-face interaction. The experimenter further

explained that the researchers were particularly interested in interactions involving members of

different ethnic groups and specified whether they were paired with a White or Aboriginal

student.

The designated White participant then received the manipulation. Those in one of the

IAT conditions completed the condition-appropriate version of the IAT, which was described as

an "information processing task." The RR-IAT followed the method of Greenwald et al. (2003)

and required participants to categorize White faces, Aboriginal faces, pleasant words (e.g.

"love"), and unpleasant (e.g. "filth") words as quickly as possible. The words and pictures were

all taken from the "Project Implicit" website. *D* scores were calculated according to Greenwald

et al.'s (2003) improved scoring algorithm. As in much IAT research, the average *D* score was significantly greater than 0, indicating pro-White bias, $M = 0.78$, $t(27) = 7.08$, $p < .001$: Participants responded more quickly when the response keys used to indicate "White" and "Aboriginal" were the same as those used to indicate "pleasant" and "unpleasant" respectively, as compared with when "White" was paired with "unpleasant" and ''Aboriginal" with "pleasant." The RI-IAT was directly parallel but involved categorizing pictures of insects and flowers (pro-flower bias, $M = 1.04$, $p < .001$). Those in the no-task control condition did neither of the tasks.

The experimenter then brought the pair members together and gave them the discussion topics (e.g., positive and negative academic and social experiences, opinions about social issues (capital punishment and euthanasia), career goals). Participants were left alone but the 15-minute discussion was audiotaped with their knowledge.

Dependent Measures

Immediately after the discussion participants completed a questionnaire, having been assured of the confidentiality of their responses. Unless otherwise noted, all ratings were made on 7-point scales. The questionnaire included four metaperception measures. First, participants completed an open-ended thought-listing task that involved listing the main thoughts that were on their mind during the discussion. Two independent raters blind to condition counted the number of positive (e.g. "I think she liked me," "he was quite interested in what I had to say") versus neutral (e.g., "I am unsure how she feels," "His feelings toward me are neutral") or negative (e.g. "she didn't seem very interested (couple of eye rolling)," "this girl was judging [me] a little") comments participants made about how the other participant viewed them. A difference score was created by subtracting the negative and neutral comments from the positive comments; square root transformations were conducted to reduce positive skew.[1] Raters'

judgments ($\alpha = .69$) were averaged together. Next, a five-item scale assessed participants' metaperceptions regarding how the other participant felt toward them (i.e., how much the other participant liked, respected, and admired them, was interested in getting to know them, and felt a strong bond with them; $\alpha = .86$). They then completed a "metaevaluation thermometer" on which they provided a number between 0 (extremely unfavorable) and 100 (extremely favorable) to indicate the other participant's overall feelings toward them. Finally, they described how they thought that the other participant felt toward them in their own words. These responses were coded in the same fashion as those to the open-ended thought-listing ($\alpha = .89$). Participants were thanked and debriefed at the end of the study.

Results

The analyses focused on the metaperceptions formed by the partners of the White participants who received the manipulation. Partners' scores across all four metaperception measures were standardized, combined together ($\alpha = .74$), and analyzed in multiple regression. The predictors were two contrast vectors comparing each of the experimental conditions to the no-task control condition (e.g., for the RR-IAT contrast, RR-IAT = 1, no-task control = -1, and RI-IAT = 0), one contrast vector representing the type of exchange (intergroup =1 and intragroup = -1), and the interactions between type of exchange and the race-relevant and race-irrelevant contrasts. Pair members' ingroup identification scores were included as covariates, along with a variable representing whether pair members shared the same first language (87% did); all covariates were centered. All significant effects involving the manipulations are reported. The predicted values and results of simple effects analyses are depicted in Table 1.

The analysis yielded a significant RR-IAT x Type of Exchange contrast, $b = -0.28$, $\beta = -.31$, $t(81) = 2.38$, $p < .025$. In line with predictions, Aboriginal partners felt that they were

viewed less positively if the White participant had done the RR-IAT than if he or she was in the no-task control condition. This was not true for White partners (indeed there was a nonsignificant trend in the opposite direction in intragroup exchanges) or when the designated White participant had done the RI-IAT.

Study 2

Study 2 was parallel to Study 1 but included only intergroup pairs and replaced the race-irrelevant IAT condition with a condition in which participants completed an explicit self-report measure of attitudes toward Aboriginal persons. This condition was included so that the effects of completing an implicit measure, the RR-IAT, could be distinguished from the effects of completing any measure referencing the ethnic group of an interaction partner in advance of an intergroup exchange.

Several potential mediators of the effect of the RR-IAT on Aboriginal partners' metaperceptions were considered. To assess whether the manipulation affected the extent to which White participants adopted a cautious and inhibited approach to the interaction, Aboriginal partners rated the White participant's degree of self-disclosure. To assess whether the manipulation affected White participants' self-efficacy, both participants made judgments about each person's relative control over the exchange. Finally, the extent to which participants activated stereotypes associated with each ethnic group was assessed.

Method

Participants

Participant characteristics were the same as in Study 1, except that each of 46 same-sex pairs of Canadian introductory psychology students (67% female) included one White and one Aboriginal student, and the perceived choice component of the self-determination scale (see

Sheldon, Ryan, & Reis, 1996) was included as an additional covariate to capture individual differences relevant to feelings of control. White pair members were randomly assigned to the RR-IAT, explicit-task, or no-task control condition.

Procedure

The procedure was identical to Study 1, except that the explicit-task condition replaced the RI-IAT condition. Participants in this condition reported their attitudes toward Aboriginal people by completing a modified version of Brigham's (1993) Attitudes toward Blacks Scale in which "Aboriginal people" replaced "black people" (e.g., "It would not bother me if my new roommate was an Aboriginal person"). Scores indicated low levels of explicit prejudice on the 16-item scale ($M = 3.44$, $SD = 1.51$ on the 10-point scale; $\alpha = .92$). Pro-White bias on the RR-IAT was again evident, $M = 0.78$, $t(14) = 6.00$, $p < .001$.[2]

Dependent Measures

This study included only one metaperception measure, the same metaperception scale used in Study 1 with one extra item assessing how much the other participant had formed a positive impression ($\alpha = .85$). A parallel impression measure assessed reactions to the other participant ($\alpha = .89$). Participants also answered questions regarding their perceptions of the other participant's specific behaviors, including perceived self-disclosure ("How much personal information about him/herself did the other participant share with you?"). Whether the impression measure appeared before or after the other items was counterbalanced. Participants also answered three questions regarding who exerted the most influence over the exchange (e.g., "The other participant/I was mostly in control of the discussion"; $\alpha = .87$); higher numbers reflected perceptions of the other participant being more in control.

Next, participants did a lexical decision-making task (presented as an information

processing task) that measured stereotype activation. The procedure and words were identical to

those used by Vorauer, Hunter, Main, and Roy (2000). Participants indicated whether each of 72

letter strings formed a word. Nine of the 36 words were relevant to White Canadians' stereotype

of Aboriginal Canadians (e.g., *lazy*), 9 were relevant to White Canadians' meta-stereotype

regarding how they are viewed by Aboriginal Canadians (e.g., *unfair*), 9 were not relevant to

either stereotype, and 9 were positive fillers. Response times to stereotype-relevant words were

trimmed so as not to exceed 2 standard deviations above the mean for any given word. Indices

were computed by averaging the response times for the traits in each category.

Results

Analyses were parallel to Study 1, with the predictors being two contrast vectors

comparing each of the experimental conditions to the no-task control condition. Covariates were

the same as in Study 1 (74% shared the same first language) but also included pair members'

self-determination scores. The predicted values are depicted in Table 2.

Metaperceptions

The analysis of Aboriginal partners' metaperceptions regarding how the White participant

viewed them yielded a significant RR-IAT contrast, $b = -0.36$, $\beta = -.38$, $t(38) = 2.13$, $p < .05$,

whereby they felt that they were viewed less positively if the White participant had done the RR-

IAT than if he or she was in the no-task control condition. There were no significant effects

involving the manipulation on Aboriginal partners' impressions of the White participant, $M = $

5.51, $SD = 0.88$.

Mediators

In line with the enhanced-caution interpretation of the RR-IAT effect, Aboriginal partners perceived that the White participant self-disclosed less if he or she had done the RR-IAT than if he or she was in the no-task control condition, RR-IAT contrast $b = -0.64$, $\beta = -.41$, $t(38) = 2.28$, $p < .05$.

In line with the reduced-efficacy account, Aboriginal partners perceived that the White participant had less control over the discussion relative to themselves if he or she had done the RR-IAT than if he or she was in the no-task control condition, RR-IAT contrast, $b = -0.46$, $\beta = -.37$, $t(38) = 2.17$, $p < .05$. This analysis also yielded a significant explicit contrast in the opposite direction, $b = 0.49$, $\beta = .39$, $t(38) = 2.45$, $p < .025$.

Analyses of the stereotype activation data included reaction times to the stereotype-irrelevant words (on which there were no significant or marginal effects) as a further covariate. Aboriginal partners showed greater activation of the Aboriginal stereotype if the White participant had done the RR-IAT than if he or she was in the no-task control condition, RR-IAT contrast, $b = -29.98$, $\beta = -.24$, $t(35) = 2.14$, $p < .05$. There were no significant effects on Aboriginal partners' activation of the White stereotype ($M = 718$, $SD = 121$ ms).

There were no significant effects on White participants' responses to the measures of any of the potential mediators. Conceivably, the manipulation interfered with White participants' post-interaction responses much as it disrupted their behavior during the interaction itself. For example, if the RR-IAT primed the possibility of bias it may have increased the influence of social desirability motivations or defensiveness on their answers. Notably, the disconnect between self-reported and outwardly observed control seems parallel to how individuals' psychological sense of power can exist outside of conscious awareness (Smith & Galinsky, 2010).

Mediation

Mediation analyses were conducted to test the various accounts for the negative effect of White participants' RR-IAT experience on their Aboriginal partner's metaperceptions. Bootstrapping procedures (Shrout & Bolger, 2002) were used to compute a confidence interval around the indirect effect (i.e., the path through the mediator), with the SPSS macros that Preacher and Hayes (2004) provide for this procedure. If zero falls outside this interval, mediation can be said to be present.

In the test of the enhanced-caution account, the RR-IAT contrast was the independent variable, Aboriginal partners' metaperceptions were the dependent variable, and Aboriginal partners' perceptions of the White participant's self-disclosure were the mediator. In this and all other mediation analyses reported, the explicit contrast was controlled along with covariates associated with significant or marginal effects. Results revealed a 95% confidence interval ranging from -0.343 to -0.004, indicating significant mediation ($p < .05$). Parallel analyses testing whether Aboriginal partners' perceptions of the White participant's relative control over the interaction (reduced-efficacy account) or Aboriginal partners' activation of the Aboriginal stereotype (enhanced-stereotyping account) accounted for the RR-IAT effect did not indicate significant mediation, CIs -0.025 to 0.207 and -0.170 to 0.127 respectively. Interestingly, however, the negative effect of the RR-IAT on Aboriginal partners' perceptions of the White participant's self-disclosure was mediated by Aboriginal partners' perceptions of the White participant's relative control over the interaction, CI -0.554 to -0.028, suggesting a link between reduced efficacy and a cautious approach to intergroup interaction (reverse mediation was not significant, CI -0.480 to 0.012).

General Discussion

The results of these studies clearly indicate that being alerted to potential bias and limited response control via doing the IAT has negative implications for White individuals' intergroup interaction behavior. In both studies, when White individuals did the RR-IAT before engaging in a face-to-face intergroup interaction, their Aboriginal interaction partner left the exchange feeling less positively regarded. No such effects were evident when White individuals completed an RI-IAT (Study 1) or an explicit prejudice measure (Study 2) before the exchange, or when their interaction partner was White (Study 1).

Mediation analyses conducted in Study 2 suggested that the experience of doing the RR-IAT led White individuals to communicate less positive regard because it prompted them to adopt a cautious approach to the interaction, as reflected by their Aboriginal partner's perceptions that their personal self-disclosures were limited. The experience of doing the RR-IAT also made White participants appear to their Aboriginal partner as relatively less efficacious and influential in the discussion, an effect that contributed to their being perceived as having shared less information about themselves during the exchange.

Limitations

Although the effects of doing the RR-IAT on White individuals' subsequent intergroup interaction behavior were clearly negative in these studies, it is possible that with a greater time delay between the test and the interaction more positive effects would be evident: With a better opportunity to reflect, introspect, and plan, individuals might work with the test experience in a positive way to modify their behavior. Further, alternative implicit tasks that frame the judgments differently (e.g., the "personalized" IAT, Olson & Fazio, 2004) might have more positive implications. Other important questions center on whether the effects might be different if taking the test was self-initiated outside of the lab and on whether the results generalize to

different intergroup relationships and to broader awareness interventions such as public

information campaigns and educational workshops. To the extent that the specters of bias and

limited control are raised in these alternative tasks and circumstances, the effects may be similar.

Along these lines, given that most people evidence bias on the test, it seems likely that the effects

would if anything tend to be stronger if feedback were provided, although this is not currently

known. Finally, the enhanced-caution account for the results was tested somewhat indirectly in

this research, via perceived self-disclosure. Although there are good grounds for expecting lower

self-disclosure to be indicative of caution and inhibition (e.g., Post, Wittmaier, & Radin, 1978;

Stephan, Stephan, Wenzel, & Cornelius, 1991), other interpretations are possible.

Implications

One direct implication of the present findings is that researchers using the IAT should be

mindful of how administering this measure might affect their participants' subsequent behavior

toward an outgroup interaction partner. More broadly, the findings suggest that being alerted to

potential bias and limited response control through a direct personal experience such as that

provided by the IAT – and possibly other interventions that seek to enhance awareness of bias

and limited control over behavioral responses – can lead to worse rather than better behavior

regulation. Indeed, the present research reveals that, at least in some circumstances, doing the

IAT can harm individuals' subsequent intergroup interaction behavior and thus suggests that

anyone who does the test should be wary of this potential unintended negative side effect.

Acknowledgements

Footnotes

1. Neutral and negative metaperceptions were combined on the basis of research indicating that

individuals experience neutral reactions from others as rejecting and as evidence of latent

negativity (Leary, Haupt, Strausser, & Chokel, 1998); examples are from both open-ended tasks.

2. Degrees of freedom on the IAT and lexical decision-making task vary due to missing data

(i.e., participant failure to respond or respond accurately; problems saving the data).

References

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse

 Jackson 'fail' the Implicit Association Test?" *Psychological Inquiry, 15*, 257-278.

Bigler, R. S., & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping:

 Theoretical and educational implications of a cognitive-based intervention. *Child

 Development, 63,* 1251–1363.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure.

 Annual Review of Sociology, 34, 277-297.

Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination

 among African–Americans: Implications for group identification and wellbeing. *Journal of

 Personality and Social Psychology, 77*, 135–149.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology,

 23*, 1933-1927.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components.

 Journal of Personality and Social Psychology, 56, 5-18.

Devine, P. G., Evett, S. R., & Vasquez-Suson, K. A. (1996). Exploring the interpersonal

 dynamics of intergroup contact. In R. Sorrentino & E. T. Higgins (Eds.), *Handbook of

 motivation and cognition: The interpersonal context (Vol. 3)*. New York: Guildford Press.

Doerr, C., Plant, E. A., Kunstman, J. W., & Buck, D. (2011). Interactions in Black and White:

 Racial differences and similarities in response to interracial interactions. *Group Processes

 and Intergroup Relations, 14*, 31-43.

Dovidio, J. F., Gaertner, S. L., Validzic, A., Matoka, K., Johnson, B., & Frazier, S. (1997). Extending the benefits of re-categorization: Evaluations, self-disclosure and helping. *Journal of Experimental Social Psychology, 33,* 401-420.

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The Race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30*, 1611-1624.

Gladwell, M. (2005). *Blink*. New York: Back Bay Books/Little, Brown and Company.

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.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Greenwald, A. G., Poehlman, Uhlmann, & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97,* 17-41.

Kaiser, C. R., & Wilkins, C. L. (2010). Group identification and prejudice: Theoretical and empirical advances and implications. *Journal of Social Issues, 66*, 461-476.

Leary, M. R., Haupt, A. L., Strausser, K. S., & Chokel, J. T. (1998). Calibrating the sociometer: The relationship between interpersonal appraisals and the state self-esteem. *Journal of Personality and Social Psychology, 74*, 1290-1299.

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*, 302-318.

Major, B., Quinton, W. J., & Schmader, T. (2003). Attributions to discrimination and self-esteem: Impact of group identification and situational ambiguity. *Journal of Experimental Social Psychology, 39*, 220-231.

Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology, 16*, 113-154.

Moss, R. (2009, June 22). We're all prejudiced and we need to admit it. *Personnel Today.* Retrieved May 2011, from, http://www.personneltoday.com/articles/2009/06/22/51170/were-all-prejudiced-and-we-need-to-admit-it.html.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36-88.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Associations Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Plant, E. A., & Butz, D. A. (2006). The causes and consequences of an avoidance-focus for interracial interactions. *Personality and Social Psychology Bulletin, 32,* 833-846.

Post, A. L., Wittmaier, B. C., & Radin, M. E. (1978). Self-disclosure as a function of state and trait anxiety. *Journal of Consulting and Clinical Psychology, 46*, 12-19.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717-731.

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming

automatic stereotype activation by mindset priming. J*ournal of Experimental Social

Psychology, 41*, 506-514.

Science Daily. (2009, June 18). *Test that found widespread unconscious racial bias validated*.

Retrieved May 2011, from,

http://www.sciencedaily.com/releases/2009/06/090617142120.htm**.**

Sheldon, K. M., Ryan, R. M., & Reis, H. (1996). What makes for a good day? Competence and

autonomy in the day and in the person. *Personality and Social Psychology Bulletin, 22,*

1270-1279.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New

procedures and recommendations. *Psychological Methods, 7*, 422-445.

Smith, P. K., & Galinsky, A. D. (2010). The nonconscious nature of power: Cues and

consequences. *Social and Personality Psychology Compass, 4*, 918-938.

Stephan, W. G., Stephan, C. W., Wenzel, B., & Cornelius, J. (1991). Intergroup interaction and

self-disclosure. *Journal of Applied Social Psychology, 21*, 1370-1378.

Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. (2000). Meta-stereotype activation:

Evidence from indirect measures for specific evaluative concerns experienced by members

of dominant groups in intergroup interaction. *Journal of Personality and Social

Psychology, 78*, 690-707.

Vorauer, J. D., & Turpie, C. (2004). Relation of prejudice to choking versus shining under

pressure in intergroup interaction: The disruptive effects of vigilance. *Journal of

Personality and Social Psychology, 87*, 384-399.

Table 1

Study 1: Predicted Values for Partners' Scores on the Metaperception Index as a Function of Exchange Type and the Designated White Participant's Experimental Condition

	White Participant's Experimental Condition		
	No-Task Control	Race-Irrelevant IAT	Race-Relevant IAT
Exchange Type			
Intragroup (White Partner)	-0.25_a (0.81)	0.02_a (0.77)	0.18_a (0.79)
Intergroup (Aboriginal Partner)	0.14_a (0.76)	0.24_a (0.45)	-0.36_b* (0.80)

Note. IAT = Implicit Association Test. Standard deviations are in parentheses. Partners' scores on the metaperception index were computed by standardizing and averaging coders' ratings of their metaperception responses on the open-ended thought-listing task, their responses to the metaperception scale, their responses to the metaevaluation thermometer, and coders' ratings of their responses to the open-ended metaperception question. Higher scores reflect more positive perceived evaluations. Significant contrasts between the experimental and no-task control conditions are marked with an asterisk ($p < .05$, two-tailed). The simple effect of exchange type within experimental condition was marginal ($p = .06$) for values not sharing a common subscript.

Table 2

Study 2: Predicted Values for Aboriginal Partners' Experience of the Exchange as a Function of the White Participant's Experimental Condition

	White Participant's Experimental Condition		
	No-Task Control	Explicit (ATA)	IAT
Metaperceptions	4.81 (0.53)	4.77 (0.84)	4.26* (0.72)
Perceptions of White Participant's Control	4.16 (1.03)	4.68* (0.75)	3.73* (0.87)
Impressions of White Participant's Self-Disclosure	5.51 (1.15)	5.77 (1.02)	4.68* (1.34)
Activation of Aboriginal Stereotype	744 (53.90)	714 (45.17)	684* (67.05)

Note. ATA = Attitudes toward Aboriginal persons scale; IAT = Implicit Association Test. Standard deviations are in parentheses. Aboriginal partners' metaperceptions, perceptions of the White participant's control, and perceptions of the White participant's self-disclosure were assessed on 7-point scales where higher numbers reflected more positive metaperceptions regarding how they were viewed, perceptions that the White participant had greater control over the discussion, and perceptions that the White participant had disclosed more. For activation of the Aboriginal stereotype, responses are in milliseconds such that lower numbers reflect greater

activation (response times to stereotype-irrelevant words are controlled). Significant contrasts

between the experimental and control conditions are marked with an asterisk ($p < .05$, two-

tailed).