Journal of Personality and Social Psychology
2000, Vol. 79, No. 6, 1022-1038

0022-3514/00/$5.00 DOI: 10.1037//0022-3514.79.6.1022

Using the Implicit Association Test to Measure Self-Esteem and Self-Concept

Anthony G. Greenwald and Shelly D. Farnham
University of Washington

Experiment 1 used the Implicit Association Test (IAT; A. G. Greenwald, D. E. McGhee, & J. L. K. Schwartz, 1998) to measure self-esteem by assessing automatic associations of self with positive or negative valence. Confirmatory factor analysis (CFA) showed that two IAT measures defined a factor that was distinct from, but weakly correlated with, a factor defined by standard explicit (self-report) measures of self-esteem. Experiment 2 tested known-groups validity of two IAT gender self-concept measures. Compared with well-established explicit measures, the IAT measures revealed triple the difference in measured masculinity–femininity between men and women. Again, CFA revealed construct divergence between implicit and explicit measures. Experiment 3 assessed the self-esteem IAT's validity in predicting cognitive reactions to success and failure. High implicit self-esteem was associated in the predicted fashion with buffering against adverse effects of failure on two of four measures.

This research developed from the assumption that distinct implicit and explicit self-esteem constructs require different measurement strategies. In particular, the research pursued implications of Greenwald and Banaji's (1995) definition of implicit self-esteem as "the introspectively unidentified (or inaccurately identified) effect of the self-attitude on evaluation of self-associated and self-dissociated objects" (p. 11).

Greenwald and Banaji's analysis summarized a widespread recent development of the view that people process social information not only in an explicit (or aware or controlled or reflective or declarative) mode but also in an implicit (i.e., unaware, automatic, intuitive, or procedural) mode (Bargh, Chaiken, Govender, & Pratto, 1992; Devine, 1989; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald & Banaji, 1995; Kihlstrom & Cantor, 1984; Wilson, Lindsey, & Schooler, 2000). The idea of implicit operation of the self has appeared in a number of recent works (Greenwald & Banaji, 1995; Hetts, Sakuma, & Pelham, 1999; Kitayama & Karasawa, 1997; Spalding & Hardin, 1999). These works, in turn, have roots in earlier research on indicators of the self's automatic operation (e.g., Bargh & Tota, 1988; Markus, 1977; Nuttin, 1985; Rogers, Kuiper, & Kirker, 1977).

The distinction between explicit and implicit operation of the self is especially interesting if it turns out that the self functions differently in these two modes. Accordingly, it is useful to be able to measure self-esteem and self-concept in ways that can distinguish the self's implicit and explicit operations. Explicit measurement of self-concept has a long history (reviewed by Wylie, 1974, and not recapitulated here), whereas there is only a much sparser history of attempts to capture the self in an implicit mode of operation. Projective measures, such as the Thematic Apperception Test (McClelland, Atkinson, Clark, & Lowell, 1953; Murray, 1943), represented the state of the art until the late 1970s, when Rogers et al. (1977) proposed the use of latencies of trait self-descriptiveness judgments in self-concept assessment.

The Rogers et al. (1977) strategy of using trait self-descriptiveness judgments was tried in numerous laboratory studies (reviewed, e.g., by Kihlstrom & Cantor, 1984; Greenwald & Pratkanis, 1984). However, the limited sensitivity of these measures to individual differences led to their being used mostly to examine aggregated effects, either in the form of preexisting differences between groups or in the form of effects of experimental manipulations. In the 1990s, there has been renewed attention to implicit measures, leading to several new procedures for assessing implicit self-concept (Aidman, 1999; Bosson, Swann, & Pennebaker, 2000; Farnham, Greenwald, & Banaji, 1999; Otten & Wentura, 1999; Pelham & Hetts, 1999; Perdue, Dovidio, Gurtman, & Tyler, 1990). This article reports the first studies that used the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) as the basis for assessing the self's implicit mode of operation.

The Implicit Association Test

The IAT (Greenwald et al., 1998) is a general-purpose procedure for measuring strengths of automatic associations between concepts. The IAT can be illustrated with a thought experiment. Imagine sorting a standard deck of 52 playing cards, containing 13 cards in each of the four suits of clubs, diamonds, hearts, and spades. You are asked to place clubs and spades in a stack to your left, and diamonds and hearts to your right. The speed with which you can do this sorting should reflect the strength of your associations within the two pairs of categories that have to be sorted together. If two suits that must be sorted together are easily associated because of some shared attribute, the task should be

Anthony G. Greenwald and Shelly D. Farnham, Department of Psychology, University of Washington.

This research was supported by National Science Foundation Grants SBR-9422242 and SBR-9710172 and by National Institute of Mental Health Grants MH-41328, MH-01533, and MH-57672.

Correspondence concerning this article should be addressed to Anthony G. Greenwald, Department of Psychology, University of Washington, Box 351525, Seattle, Washington 98195-1525. Electronic mail may be sent to agg@u.washington.edu.

relatively simple. In this example, shared color attributes provide a basis for association that makes it easy to sort clubs and spades (shared attribute: black color) to the left and hearts and diamonds (red) to the right.

What happens if color cannot be used as a grouping attribute? If your task is to sort clubs and diamonds to the left and spades and hearts to the right, then the black-left, red-right strategy no longer works, and your sorting speed should deteriorate. Interestingly, this second sorting task should discriminate bridge players from nonplayers. For bridge players, hearts and spades are well associated because they are the higher ranking suits in that game. Any bridge player can readily observe the effect of these suit–rank associations by trying to do both a rank-consistent sort (clubs+diamonds vs. hearts+spades) and a rank-inconsistent sort (clubs+hearts vs. spades+diamonds)—the rank-consistent sort should be noticeably faster for players who have learned suit ranks in a game such as bridge, but others should not show a similar speed difference.

Described abstractly, the IAT's procedure has the subject give one response to two sets of items that represent a possibly associated concept–attribute pair and a different response to a second pair of item sets that is selected to complement the first two. Association between the concept and attribute that share a response is inferred to be stronger the faster the subject performs the task. In the first investigation of the IAT, Greenwald et al. (1998) asked subjects to sort each of a series of computer-presented words by rapidly pressing a left-side or right-side key on a computer keyboard. The automatic association between a concept (e.g., flowers) and an attribute (e.g., positive valence) was measured by observing the difference in speed between a condition in which flower names and pleasant-meaning words shared the same response key (this was typically fast) and a condition in which flower names and unpleasant-meaning words shared the same response key (typically slow). In that experiment, the two concepts were flower and insect, and the two attributes was pleasant and unpleasant. The resulting IAT measure compared the aggregate association strength of flower–pleasant and insect–unpleasant with that of flower–unpleasant and insect–pleasant. The results indicated that, in aggregate, flower-pleasant and insect-unpleasant were stronger associations than flower-unpleasant and insect-pleasant.

Because it uses complementary pairs of concepts and attributes, the IAT is limited to measuring the relative strengths of pairs of associations rather than absolute strengths of single associations. In practice, however, the IAT can nevertheless be effectively used because many socially significant categories form complementary pairs, such as positive–negative (valence), self–other, male–female, Jewish–Christian, young–old, weak–strong, warm–cold, liberal–conservative, aggressive–peaceful, etc.[1]

The IAT was readily adapted to measuring implicit self-concept by observing response speeds for classification tasks in which the concept pair used in the IAT was self–other. Thus, the self-esteem IAT introduced in Experiment 1 compared self–pleasant and other–unpleasant associations with self–unpleasant and other–pleasant. Similarly, the gender self-concept IAT introduced in Experiment 2 compared self–feminine and other–masculine associations with self–masculine and other–feminine.

Validity of the IAT

The first investigations of the IAT (Greenwald et al., 1998) confirmed that the IAT could detect valence differences that were associated both with familiar nonsocial objects (flowers, musical instruments, insects, and weapons) and with significant social objects (Japanese and Korean ethnicity, and Black and White race). Greenwald et al. (1998) also demonstrated that IAT measures were stable across several procedural variations, including whether the pleasant category was assigned to a left-side or right-side response, the time interval between response to one stimulus and presentation of the next stimulus item (varied from 150 to 750 ms), and whether concepts and attributes were represented by 5 or 25 items. Observed IAT effects were also quite stable over variations in the manner of treating data from error responses and in the strategies used to deal with the typically skewed (extended upper tail) latency distributions. Later research provided additional internal validity evidence, establishing that the IAT's association-strength measure was not influenced by variations in familiarity of items used to represent the contrasted concepts (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Ottaway, Hayden, & Oakes, in press; Rudman, Greenwald, Mellott, & Schwartz, 1999).

Several researchers have demonstrated that IAT measures can be influenced in theoretically expected fashion by procedures that might be expected to influence automatic attitudes or stereotypes. Dasgupta and Greenwald (2000) showed that viewing photos of admirable members of stigmatized groups (African Americans or elderly) and despised members of nonstigmatized groups (European Americans or young) reduced automatic negative associations toward those groups. Blair and Ma (1999) found that writing an imagined description of a strong woman decreased IAT-measured association of male (more than female) with strength. And Rudman, Ashmore, and Gary (1999) reported that an IAT measure of race preference for White over Black was reduced among students who had completed a Prejudice and Conflict seminar taught by an African American instructor (see also Lowery & Hardin, 1999).

Going beyond sensitivity to age (Mellott & Greenwald, 1999), gender (Rudman, Greenwald, & McGhee, in press), and racial and ethnic (Greenwald et al., 1998) group differences, the IAT has also been shown to be sensitive to individual differences. Correlations between parallel IAT measures of various attitudes were reported in Greenwald et al.'s (1998) Experiments 2 (r = .85) and 3 (r = .46), and by Dasgupta et al. (2000; r = .39). Test–retest reliabilities of r = .65 and r = .69 were reported, respectively, by Dasgupta and Greenwald (2000), and by Bosson et al. (2000). Their variability notwithstanding, these figures average to indicate moderately good stability ($\bar{r}$ = .64, using r-to-Z method). Theoretically meaningful correlations of IAT measures of ingroup favoritism with multiple indicators of degree of ingroup identity as

[1] There have been some moderately successful attempts to use the IAT in designs that allow comparison of just two associations involving just one target concept, rather than comparison of two pairs of associations involving two complementary target concepts. For example, Swanson, Rudman, and Greenwald (in press) compared the association of a single target concept, smoking, with pleasant versus pleasant. Nosek and Banaji (2000) have developed a more general approach to a one-category IAT measure and have demonstrated its use with both nonsocial and social objects.

Japanese or Korean were reported by Greenwald et al. (1998, Experiment 2). Rudman and Glick (1999) reported a correlation between prejudice against female applicants in a simulated job interview and IAT-assessed gender stereotypes. Convergent validity with alternative latency-based measures of implicit attitudes has been demonstrated in correlations of IAT measures with semantic priming measures of association strength (Cunningham, Preacher, & Banaji, in press; Mellott & Greenwald, 2000; Rudman & Kilianski, 2000). And convergence of IAT-measured automatic race preferences with a physiological measure (fMRI-measured amygdala activation of White participants while viewing unfamiliar African American faces) has been reported by Phelps et al. (2000).

Measuring Implicit Self-Esteem With the IAT

The self-esteem IAT involves five steps (see Figure 1). In each step, the subject presses a left or right key to rapidly categorize each of a series of stimuli that are presented in the middle of a computer screen. Instructions for the categorization task vary for the five steps, and latency is measured and averaged for each task variation. In the first step, subjects practice a *target concept* discrimination by categorizing items into *self* and *other* categories. In the second step, subjects practice an *attribute* discrimination by categorizing items into *pleasant* and *unpleasant* categories. Third, subjects categorize items into two combined categories, each including the target and attribute concept that were assigned to the same key in the preceding two steps (e.g., self+pleasant for the left key and other+unpleasant for the right key). The fourth step provides practice that reverses key assignments for either the target or attribute concept. Finally, the fifth step is like the third, but it uses the just-switched key assignments (e.g., self+unpleasant to the left, and other+pleasant to the right). Implicit self-esteem is measured in the form of an IAT effect, computed as the difference in mean latency between Steps 3 and 5. The self-esteem IAT effect measures how much easier it is for subjects to categorize self items with pleasant items than self items with unpleasant items. Half of the subjects do the sequence of five tasks interchanging the positions of Steps 2 and 3 with Steps 4 and 5 to counterbalance possible task order effects (Greenwald et al., 1998).

	Category labels	Sample items	Category labels
Step 1: practice block (20 trials)	not me		me
	○	self	●
	●	other	○
Step 2: practice block (20 trials)	unpleasant		pleasant
	○	joy	●
	●	vomit	○
Step 3: practice block (20 trials) critical block (40 trials)	unpleasant or not me		pleasant or me
	○	self	●
	○	joy	●
	●	other	○
	●	vomit	○
Step 4: practice block (20 trials)	pleasant		unpleasant
	●	joy	○
	○	vomit	●
Step 5: practice block (20 trials) critical block (40 trials)	unpleasant or me		pleasant or not me
	●	self	○
	○	joy	●
	○	other	●
	●	vomit	○

Figure 1. Categorization tasks for the five steps of the self-esteem Implicit Association Test (IAT). Black dots indicate the correct response. The IAT effect is the difference in response times between Steps 3 and 5. The orders of Steps 2–3 and Steps 4–5 were counterbalanced because of possible effects of having the self+pleasant versus the self+unpleasant combination first.

Experiment 1: Implicit and Explicit Self-Esteem Compared

Experiment 1 was the first experiment to use the IAT to measure an aspect of self-concept—in particular, it measured implicit self-esteem. An obvious initial question to ask of a measure of implicit self-esteem is how it relates to existing self-report (or explicit) measures of self-esteem. There are two reasons to expect convergence between measures of implicit and explicit self-esteem. First, in responding to self-report measures of self-esteem, subjects presumably attempt to introspectively access their association of self with positive valence, which is what the implicit self-esteem IAT seeks to measure. Second, in repeatedly expressing positive self views on explicit measures, subjects practice and presumably strengthen the association of self with positive valence (cf., Fazio, Powell, & Herr, 1983).

At the same time, implicit and explicit self-esteem may not be strongly related because several known influences on responses to self-report measures could affect implicit measures differently, less, or not at all. These influences on self-report measures include demand characteristics (Orne, 1962), evaluation apprehension (Rosenberg, 1969), impression management (Tedeschi, Schlenker, & Bonoma, 1971), self-deception (Gur & Sackeim, 1979), and self-enhancement (Greenwald, 1980; Taylor & Brown, 1984).

Experiment 1 used confirmatory factor analysis to test whether implicit and explicit self-esteem measures (a) converged on a single construct or, alternately, (b) identified distinguishable constructs. Experiment 1 also included self-report measures of impression management and self-deception (Paulhus, 1991), in the hope that these might shed light on possible differences between implicit and explicit measures that could be due to more socially desirable responding on the explicit measures.

Method

Subjects

Students from introductory psychology courses at University of Washington participated in exchange for an optional course credit. Six subjects'

data were discarded for having error rates on the IAT in excess of 20%, suggesting that they either misunderstood instructions or were trying to respond too rapidly. One subject's data were discarded for having mean latencies over 2 s, and another subject's data were discarded for not following instructions. Additionally, five subjects were dropped for having incomplete data.[2] There remained 145 subjects, 93 female (64 Caucasian, 26 Asian, 3 Other) and 51 male (22 Caucasian, 26 Asian, 3 Other), in addition to one who declined to report sex.

Procedures

After being seated in a small room with a desktop computer, subjects first completed paper-and-pencil self-report questionnaires that assessed self-esteem, impression management, and self-deception.[3] Subjects were instructed to place their finished questionnaires in a sealed box marked "completed questionnaires," which was provided to reinforce prior instructions that subjects' anonymity and privacy were being protected. After subjects completed these questionnaires, the experimenter introduced the subject to the IAT computer program and then left the subject to complete the program in privacy. The two computer-administered IAT measures both assessed self-esteem, one assessing the associations of self versus other with pleasant- and unpleasant-meaning words and one assessing the associations of self versus other with positive and negative traits.

Explicit Measures

At the beginning of the experimental session, after subjects provided self-descriptive demographic information for age, sex, and race, they completed six self-report measures. Four of these were self-esteem measures: the Rosenberg Self-Esteem Scale (Rosenberg, 1965), the Self-Attributes Questionnaire, (SAQ; Pelham & Swann, 1989), a thermometer scale on which participants indicated how warmly they felt toward themselves on a vertical scale anchored at bottom and top by 0 and 99, and a semantic differential scale of five items that requested self-descriptions by checking one of 7 points on scales anchored at ends by bipolar adjective pairs: ugly/beautiful, bad/good, unpleasant/pleasant, dishonest/honest, and awful/nice. The Balanced Index of Desirable Responding (BIDR; Paulhus, 1991) was used to measure impression management and self-deception. The order of the six measures was counterbalanced by giving half the subjects the reverse of the order just described.

IAT Procedures

The two self-esteem IATs were administered on PC-type computers with a program that constructed idiographic IAT self-concept measures for each subject by eliciting from each a series of 18 self-descriptive (*me*) and 18 not-self-descriptive (*not-me*) items (Farnham, 1998). The two IATs assessed *affective* and *evaluative* implicit self-esteem by using, respectively, (a) pleasant and unpleasant words (e.g., diamond, health, sunrise; agony, filth, poison) as the items for the positive and negative affective concepts, and (b) positive and negative trait words (e.g., bright, noble, honest; ugly, vile, guilty) for the positive and negative evaluative concepts. The two IATs were administered in counterbalanced order. Also, for each IAT, whether the self+positive critical block was encountered first or second was counterbalanced. Complete lists of the IAT items are given in the Appendix.

To assure their understanding of the IAT procedure, subjects first completed a short tutorial that used categories unrelated to self (red vs. white colored objects and snakes vs. birds). After the tutorial, each of the two IATs consisted of seven blocks of categorization trials, with 20 trials for practice blocks and 40 trials for data-collection blocks (see Figure 1). Each stimulus item was displayed until its correct response was made. The next stimulus item then followed after a 150-ms intertrial interval. The computer recorded elapsed time between the start of each stimulus word's presentation and occurrence of the correct keyboard response.

To encourage subjects to respond rapidly while making relatively few errors, the computer displayed mean latencies in milliseconds and error rates in percent after each block. All blocks were practice blocks except for the two critical blocks from which data were used to calculate the IAT effect. The IAT effect for implicit self-esteem was computed by subtracting the mean latency for the me+positive block from that for the me+negative block (Step 5 − Step 3 in Figure 1). During data-collection blocks, stimulus items were drawn alternately from the me or not-me lists (odd-numbered trials) and from the positive or negative lists (even-numbered trials). Items from each category pair were selected randomly and without replacement so that all items were used once before any items were reused.

IAT Items

Idiographic items. Before completing the IAT, each subject provided 18 me and 18 not-me items. Me items included first and last names, hometown, phone number, birth month, and birth year (see Appendix). These items presumably did not have positive or negative qualities apart from those that might have been gained by association with self. For not-me items, subjects were instructed to pick from lists of items comparable to the me items and to select items such that chosen not-me items were (a) familiar, (b) not self-identified, and (c) neither strongly liked nor disliked. After choosing these items, subjects viewed their resulting me and not-me lists and were asked to delete items that (in retrospect) seemed inappropriate or were misspelled. Subjects were allowed to delete up to eight items from each, leaving a minimum of ten per list.

Positive and negative affective and evaluative items. Pleasant and unpleasant words were selected from the pleasantness-judgment norms of Bellezza, Greenwald, and Banaji (1986). Subjects were allowed to delete items from each list that they did not regard as pleasant or unpleasant, respectively. Positive and negative evaluative items (traits) were selected mostly from trait words that have been used in self-esteem questionnaires to represent high and low self-esteem, respectively. Subjects again had the opportunity to delete traits from each list that they did not regard as desirable or undesirable, respectively. Subjects could delete up to four items from each list, leaving a minimum of ten per list (see Appendix).

Data Reduction

IAT data for analyses were obtained only from the 40-trial data-collection blocks of Steps 3 and 5 (see Figure 1). Consistent with procedures introduced by Greenwald et al. (1998), (a) the first two trials of each data-collection block were dropped because of their typically lengthened latencies; (b) a logarithm transformation was used to normalize the distribution of latencies; (c) prior to this transformation, latencies greater than 3,000 ms were recoded to 3,000 ms, and latencies less than 300 ms were recoded to 300 ms. Alternative treatments of outlying trials, such as using different boundaries to identify outliers, excluding them entirely, or even keeping them in the data set, had no substantial impact beyond sometimes adding noise to the findings. As previously noted, subjects

[2] None of the results of analyses to be reported would be altered by including the partial data of subjects who were dropped for having incomplete data. These subjects were dropped so that Table 1's correlations would suffice to reproduce the present covariance structure analyses.

[3] The procedure of completing self-report before IAT measures was used in the present series of experiments because it was suspected that completing the IAT measures first might be more likely to influence the self-report measures than vice versa. In other studies in the authors' laboratory, no systematic effects of the order of implicit and explicit measures have been observed.

whose error rates for data-collection blocks of the IAT exceeded 20% (6 subjects) or who had mean latencies in excess of 2,000 ms (1 subject) were not included in analyses.

Results

Implicit Self-Esteem

Initial analyses examined the effects of counterbalancing variables (order of the two IATs and order of the two data-collection blocks) on IAT effects. Neither counterbalancing variable had a significant effect. Overall, subjects responded much more rapidly when associating self with positive items (see Figure 2). IAT effects (mean latency for the self+negative block minus mean latency for the self+positive block) were strong for both the affective IAT, Cohen's $\mathbf{d} = 1.38$, $F(1, 141) = 617$, $p = 10^{-53}$, and for the evaluative IAT, $\mathbf{d} = 1.46$, $F(1, 141) = 468$, $p = 10^{-46}$. The mean IAT effects for the affective and evaluative measures did not significantly differ, $F(1, 141) = 0.02$, $p = .89$. Supplementary analyses indicated that neither sex nor race moderated magnitude of either of the self-esteem IATs, all Fs < 1.

Explicit Self-Esteem

Means for the explicit self-esteem measures are reported in Table 1, classified by race and sex. Race had small effects on explicit self-esteem measures, such that Caucasians and men tended to report higher self-esteem than Asians and women. Analysis of variance (ANOVA) of an average of standardized scores for the four self-esteem measures revealed that the effect of race was statistically significant, $\mathbf{d} = .36$, $F(1, 134) = 4.73$, $p = .03$, whereas the effect of sex was not significant, $\mathbf{d} = .04$, $F(1, 134) = 0.38$, $p = .54$.

Relationships Between Measures of Implicit and Explicit Self-Esteem

Table 2 provides correlations among all of Experiment 1's measures. The two measures of implicit self-esteem were positively and significantly correlated with each other ($r = .43$), at almost the same level that the four measures of explicit self-esteem correlated with each other (average $r = .46$). Measures of implicit self-esteem had typically weak correlations with measures of explicit self-esteem. However, all eight correlations were numerically positive, and five of the eight were statistically significant (average $r = .17$). With the one exception of its positive correlation with the semantic differential self-esteem measure, the BIDR measure of impression management had near nil correlations with both implicit self-esteem (average $r = .06$) and explicit self-esteem (average $r = .08$). The BIDR measure of self-deception functioned very similarly to the explicit self-esteem measures, correlating an average of $r = .24$ with the two implicit self-esteem measures, and an average of $r = .39$ with the four explicit self-esteem measures.

Using maximum likelihood analysis, two confirmatory factor analyses (CFAs) were computed, to determine whether the data for the implicit and explicit self-esteem measures were better fit by a model with two factors (implicit and explicit self-esteem) or one

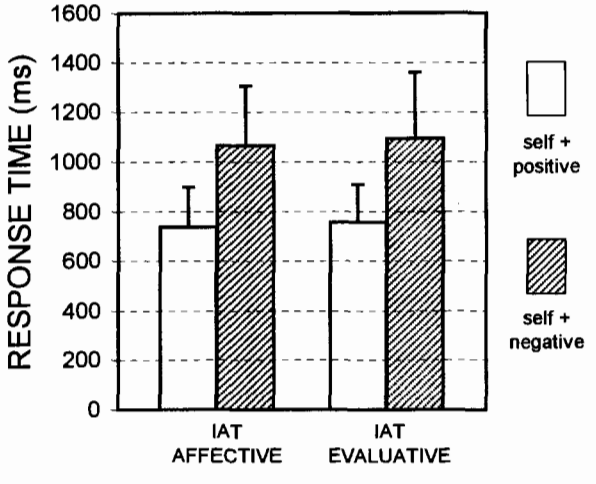

Figure 2. Response times for critical blocks for the two Implicit Association Tests (IATs) in Experiment 1. The mean IAT effect for each test is the mean for its self+positive condition subtracted from that for its self+negative condition. Error bars are standard deviations. $N = 145$.

Table 1
Experiment 1's Eight Measures Classified by Race and Sex of Subjects

Measure	Caucasian		Asian	
	Male ($n = 22$)	Female ($n = 64$)	Male ($n = 26$)	Female ($n = 26$)
Implicit self-esteem				
IAT affect (ms)				
M	291	348	313	319
SD	270	172	191	149
IAT trait (ms)				
M	329	351	327	286
SD	225	206	195	172
Explicit self-esteem				
Rosenberg SES				
M	24.32	24.33	22.42	22.81
SD	6.43	5.40	5.21	6.65
SAQ[a,b]				
M	4.47	3.64	3.22	3.01
SD	1.58	1.27	1.23	0.98
Semantic differential				
M	29.2	29.0	28.9	28.3
SD	4.92	4.46	3.59	3.96
Thermometer				
M	78	78	75	77
SD	21	14	14	15
Socially desirable responding (BIDR)				
Impression management				
M	4.50	4.86	3.50	4.62
SD	4.02	3.47	2.69	3.36
Self-deception[a]				
M	6.59	5.91	4.31	4.62
SD	3.51	3.19	2.88	2.90

Note. Total $N = 138$. The table does not include Experiment 1's 7 subjects who could not be placed in the table's demographic categories. Implicit Association Test (IAT) measures are in milliseconds. Ranges of other measures: Rosenberg Self Esteem Scale (SES), 0–30; Self-Attributes Questionnaire (SAQ), 1–10; Semantic differential, 5–35; Thermometer, 0–99; Balanced Index of Desirable Responding (BIDR) impression management and self-deception, 0–20 (with scores of 7 or greater considered to be high—Paulhus, 1991).
[a] $p < .005$ for main effect of race. [b] $p < .05$ for main effect of sex.

Table 2
Correlations Among All Measures of Experiment 1

Measure	1	2	3	4	5	6	7	8
Implicit self-esteem								
1. IAT: self-affect	—	**.432**	.130	.273	*.178*	*.198*	.005	*.198*
2. IAT: self-evaluation		—	.105	*.197*	*.201*	.105	.122	.274
Explicit self-esteem								
3. Rosenberg SES			—	**.407**	**.448**	**.738**	.023	**.474**
4. SAQ				—	*.176*	**.349**	−.001	**.345**
5. Semantic differential					—	**.524**	**.288**	**.327**
6. Thermometer						—	.018	**.412**
BIDR								
7. Impression management							—	**.293**
8. Self-deception								—

Note. $N = 145$. Decimal points omitted. IAT = Implicit Association Test; SES = Self-Esteem Scale; SAQ = Self-Attributes Questionnaire; BIDR = Balanced Index of Desirable Responding. For $N = 145$, rs of .163, .232, .286, and .331 are associated, respectively, with two-tailed p values of .05, .005, **.0005**, and **.00005**.

(self-esteem). The first analysis (Figure 3, left panel) used all six of Experiment 1's self-esteem measures. A model in which implicit and explicit self-esteem were defined as separate, but correlated, factors fit the data well, χ^2 (8, $N = 145$) = 17.40 ($p = .03$), CFI = .96, RMSEA = .09, p(close fit) = .11.[4] These statistics indicate a moderately good but not close fit (see footnote 4). By contrast, a model that constrained all six measures to represent a single factor showed a noticeably poorer fit, $\chi^2(9, N = 145) = 35.30$ ($p = .00005$), CFI = .88, RMSEA = .14, p(close fit) = .001. Because these two models are nested, a chi-square test of significance for the difference between them was possible. This test revealed that the two-factor model had significantly better fit than the one-factor model, $\chi^2(1, N = 145) = 17.90$ ($p = 10^{-5}$). In the two-factor model, the estimated correlation between the implicit and explicit self-esteem factors was .28 (see Figure 3, left panel).

The less-than-close fit of the two-factor, six-variable model led to conducting an additional set of analyses with 2 four-variable models. The four-variable models omitted the two explicit measures that were more weakly connected to the explicit self-esteem latent variable of the six-variable model. For the four-variable analysis, a model in which implicit and explicit self-esteem were defined as separate, but correlated, factors (Figure 3, right panel) fit the data extremely well, $\chi^2(2, N = 145) = 0.55$ ($p = .76$), CFI = 1.00, RMSEA = .00, p(close fit) = .82. These statistics indicate close fit (see footnote 4). In comparison, the model that constrained all four self-esteem measures to represent a single factor had noticeably poorer fit, $\chi^2(2, N = 145) = 21.70$ ($p = 10^{-6}$), CFI = .86, RMSEA = .26, p(close fit) = .0002. In the two-factor model, the estimated correlation between the implicit and explicit self-esteem factors was .22.[5]

Discussion

Both IAT self-esteem measures showed, on average, strong self-positivity. In the self+positive conditions, subjects categorized items an average of 323 ms faster than in the self+negative conditions. The two IAT self-esteem measures were positively, but weakly, correlated with explicit measures of self-esteem (average $r = .17$). Two confirmatory factor analyses were consistent in interpreting implicit and explicit self-esteem as distinct constructs that are positively, but weakly, correlated.

There were no effects of subject sex or race (Asian vs. Caucasian) on the measures of implicit self-esteem, but there was a small effect of race (Caucasians higher than Asians) on a combined index of explicit self-esteem. There was a statistically significant effect of sex (men higher than women) on one of the four explicit self-esteem measures (SAQ), but an unweighted average of the four explicit self-esteem measures did not show a statistically significant sex effect.

A possible explanation for the race effect on explicit, but not implicit, self-esteem is that Asian Americans may present themselves modestly on self-report measures. Because such a possibility was anticipated, the BIDR measure of impression management was included in Experiment 1. However, the BIDR impression management measure was essentially uncorrelated with either implicit or explicit self-esteem (see Table 2) and showed no differences as a function of race (see Table 1). Accordingly, the BIDR did not shed light on the observed higher level of explicit self-esteem for Caucasian than Asian subjects.

An additional exploratory examination of race and sex differences in the implicit and explicit self-esteem measures showed that the explicit–implicit correlation was higher for Caucasian men, $r = .51$, $p = .02$, than for the other three race-sex combinations, respectively rs = .06, .23, and .20, for Asian women, Caucasian women, and Caucasian men (all nonsignificant). Again, a self-presentational interpretation of this difference in correlation magnitudes was suspected, but, again, lack of correlations of the BIDR impression management measure with the self-esteem measures provided no support for such an interpretation. Nevertheless, the observed greater explicit–implicit correlation for Caucasian men was intriguing enough to suggest that it would be worth examining in other data collections that provide the opportunity.

Experiment 2: Validating a Measure of Implicit Gender Self-Concept

Experiment 2 provided the initial use of the IAT to measure self-concept. This was done by replacing the valence attribute used in Experiment 1 with the nonvalence attribute of masculinity–femininity. As a second variation from Experiment 1, in addition to using Experiment 1's idiographic method of representing self with subject-specific items (first name, home town, etc.) Experiment 2 also used a generic format in which self was represented by first-person pronouns (I, me, my, mine, and self).

[4] RMSEA is the root-mean-square error of approximation fit index that has been described by Browne and Cudeck (1993) and MacCallum, Browne, and Sugawara (1996). These authors characterize RMSEA < .05 as indicating close fit, .05–.08 as close-to-fair fit, .08–1.0 as mediocre fit, and RMSEA > .10 as poor fit. The p(close fit) statistic is like a null-hypothesis testing p value, such that values less than (say) .05 indicate rejection of the hypothesis of close fit.

[5] Although the one-factor and two-factor models in the four-variable analysis have a nested relationship, a 1-df test for significance for their difference was not appropriate because the two-factor model had an extra degree of freedom, which was added because of a constraint imposed in the computational routine to keep estimated error variances nonnegative.

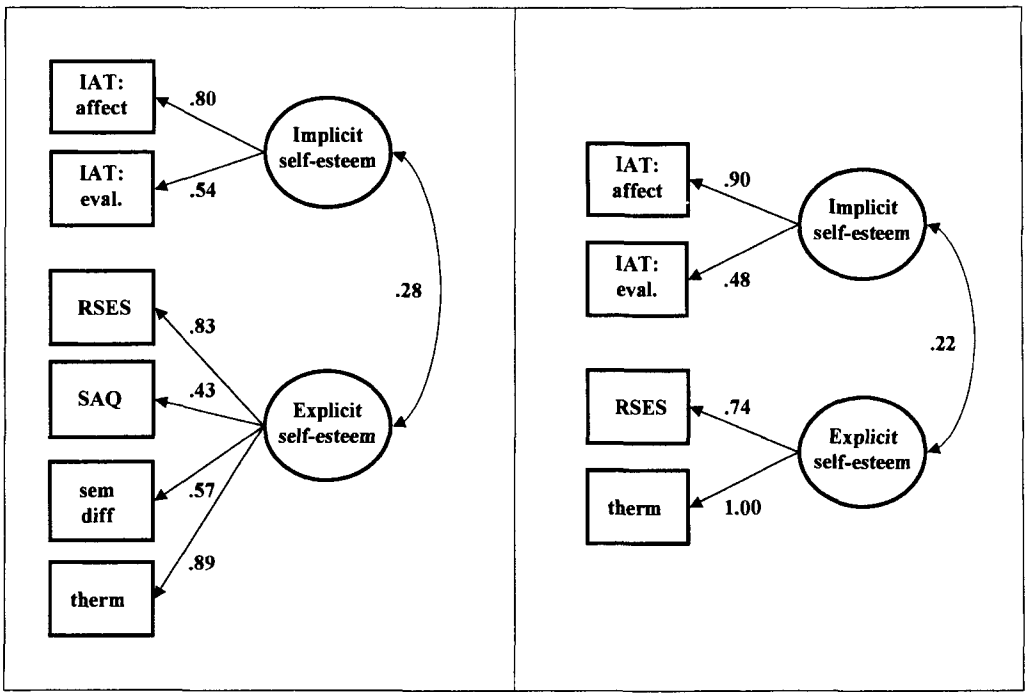

Figure 3. Confirmatory factor analyses of Experiment 1. Applying the RMSEA criterion (see footnote 4), the six-variable, two-factor model in the left panel had mediocre fit. The reduced four-variable, two-factor model on the right had close fit. Both models indicated the superiority of models with separate implicit and explicit factors to single-factor models. (Data from Experiment 1; $N = 145$.) RSES = Rosenberg Self-Esteem Scale; SAQ = Self Attributes Questionnaire; sem diff = semantic differential; therm = thermometer; IAT eval. = Implicit Association Test with trait items.

The selection of masculinity–femininity for investigation in Experiment 2 afforded a known-groups validation strategy, to see if the IAT would be sensitive to gender self-concept differences between men and women that have been identified in previous research. For example, previous research has shown that men and women differ in self-ascription of traits measured by the Bem Sex Role Inventory (BSRI; Bem, 1974) and the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974). Although these traits are commonly referred to as masculinity and femininity, they are also sometimes identified as instrumentality (more characteristic of men's self-concepts) and expressiveness (more characteristic of women's self-concepts—cf. Spence & Helmreich, 1979). If the IAT validly assesses self-concept, it should detect the previously observed difference between men and women in these aspects of self-concept. It would do this by showing that women more strongly associate self with feminine or expressive attributes than with masculine or instrumental attributes, whereas men should show the reverse pattern.

Method

Subjects

Students from introductory psychology courses at University of Washington participated in exchange for an optional course credit. Five of 66 subjects' data were discarded for having error rates in excess of 20%, and another three were dropped for having incomplete data, leaving a total of 58 (30 women, 28 men).

Procedure

Procedures were similar to those of Experiment 1. After being seated at a desktop computer in a small room, each subject first completed a set of three self-report measures of masculinity and femininity and then proceeded to the computer-administered IATs.

Questionnaires. Subjects first provided some demographic information (age and sex), and then completed the BSRI, the PAQ, and a semantic differential measure that used each of the IAT's six masculine and six feminine items (see Appendix) to define one pole of a 7-point bipolar scale. The 12 resulting scales were rough/gentle, competitive/cooperative, cold/warm, independent/dependent, tough/tender, forceful/passive, insensitive/sensitive, strong/weak, unsympathetic/sympathetic, confident/hesitant, hard/soft, and aggressive/peaceful. The averaged response to these 12 items (scored from −3 to +3, with the second item of each pair at the high end) provided a bipolar semantic differential measure of masculinity–femininity. The PAQ provided both a bipolar measure and separate unipolar measures of masculinity and femininity, whereas the BSRI provided unipolar measures of each of masculinity and femininity.

IAT measures. There were two gender self-concept IAT measures. The idiographic gender self-concept IAT differed from Experiment 1's implicit self-esteem IAT by using six masculine (e.g., rough, competitive) and six feminine (e.g., gentle, cooperative) items in place of Experiment 1's pleasant and unpleasant items. The generic gender self-concept IAT differed from the just-described idiographic version by replacing subject-generated items for me and not-me with sets of five pronouns that were used identically for all subjects. Two procedural variables for the IAT measures were counterbalanced: first, whether subjects completed the generic measure or the idiographic measure first, and second, whether the

Table 3
Implicit and Explicit Measures of Gender Identity (Experiment 2)

Measure	Men (N = 28)		Women (N = 30)		Women−men difference		
	M	SD	M	SD	Cohen's **d**	t	p
Implicit gender identity							
IAT effect—Idiographic	−104	176	172	189	1.51	6.17	10^{-7}
IAT effect—Generic	−15	150	189	125	1.49	5.77	10^{-7}
Explicit gender identity							
PAQ (bipolar)	16.4	4.24	18.7	4.14	0.57	2.16	.04
Semantic differential	0.39	0.70	0.64	0.60	0.39	1.47	.15
BSRI femininity	4.65	0.63	5.03	0.52	0.66	2.51	.01
PAQ femininity	23.3	4.34	24.2	3.06	0.24	0.89	.38
BSRI masculinity	4.83	0.91	4.54	0.87	−0.33	−1.24	.22
PAQ masculinity	20.5	5.03	18.8	4.51	−0.34	−1.30	.20

Note. IAT = Implicit Association Test; PAQ = Personal Attributes Questionnaire; BSRI = Bem Sex Role Inventory. The bipolar measures (first four in the table) are scored so that higher numbers indicate greater identity as feminine. Ranges: BSRI, 1–7; PAQ, 8–40; semantic differential, −3–+3.

self+masculine combination or the self+feminine combination was encountered first in each IAT.

Idiographic me/not-me items. Before completing the idiographic self-concept IAT, each participant provided six me and six not-me items to the computer program: first name, middle name, last name, home city, state, and country. To avoid confounding the self–other contrast with a male–female contrast, first and middle names for the not-me category were constrained to be female for female subjects and male for male subjects. Subjects had the opportunity to delete one item from each of the me and not-me lists; home country was dropped to reduce the list to five if the subject opted to keep all six items.

Masculine and feminine items. Masculine and feminine IAT items were gleaned from the BSRI and PAQ questionnaires. These items are listed in the Appendix.

Pronoun me/not-me items. Five pronouns represented the me category (I, me, my, mine, self) and another five represented not-me (they, them, their, theirs, other).

Results and Discussion

Experiment 2 was designed to compare the ability of implicit and explicit measures to detect the expected difference between men and women in masculinity–femininity of self-concept. As can be seen in Table 3, both of the implicit measures, together with all six of the explicit measures, directionally showed expected differences between men and women. Women were generally more toward the feminine end, and men were more toward the masculine end, of all eight gender self-concept measures summarized in Table 3. Nevertheless, the IAT measures had substantially larger effect sizes for this sex difference. That is, the feminine self-concepts of women and the masculine self-concepts of men were more strongly evident on IAT measures than on self-report measures. The average effect size for the two IAT measures was $\bar{\mathbf{d}}$ = 1.50, more than triple the average effect size for the six self-report measures ($\bar{\mathbf{d}}$ = 0.42).

Table 4 reports the correlations among Experiment 2's implicit and explicit measures. Additionally, a bipolar version of the BSRI was added to the data set (by subtracting the BSRI masculinity score from the BSRI femininity score), after observing that this index correlated highly with both the bipolar PAQ and the semantic differential measure. As a summary of Table 4, the correlation between the two IAT measures was r = .68, the average correlation among the three bipolar explicit measures was r = .80, and the average correlation between the two implicit and the three bipolar explicit measures was r = .32.

The correlational data just summarized suggested construct divergence between implicit and explicit measures of gender self-concepts. Figure 4 summarizes a CFA that included the two IAT measures and the three bipolar explicit measures. This CFA was designed to determine whether the IAT and explicit measures were better interpreted as defining a single construct or as defining two distinct constructs. For the two-factor model shown in Figure 4, the fit statistics were moderately good, $\chi^2(5, N = 58) = 6.60$ (p = .25), CFI = .99, RMSEA = .075, p(close fit) = .27. This two-factor model's fit was in the close-to-fair category (see footnote 4). In comparison, the one-factor model that constrained all five measures to represent a single factor had a poor fit, $\chi^2(5, N = 58) = 13.93$ (p = .02), CFI = .95, RMSEA = .18, p(close fit) = .03.[6]

Experiment 2 demonstrated that IAT's gender self-concept measures were sensitive to expected differences between men and women. There was no reason to anticipate the finding that IAT measures would show greater sex differences than did the several explicit measures. A possible interpretation is that the self-report measures may be susceptible to self-presentation strategies that are different for men and women. Perhaps these differences can be understood by considering recent shifts in (at least, American) views of ideal women and men. The ideal American woman is now seen as more assertive than the ideal American woman of previous generations, and the ideal man is now seen as more sensitive than

[6] The two-factor model required a computational constraint to assure that error variances were nonnegative. This computational constraint added a degree of freedom and precluded a test of significance for the difference between the otherwise nested models. It is nevertheless apparent from the fit statistics that the one-factor model has substantially poorer fit than the two-factor model.

Table 4
Correlations Among All Measures of Experiment 2

Measure	1	2	3	4	5	6	7	8	9	10
1. Subject sex	—	**636**	**611**	*277*	193	*309*	*318*	119	−164	−171
Implicit bipolar gender self-concept										
2. IAT: idiographic items		—	**677**	**454**	*353*	398	*270*	201	*−304*	*−279*
3. IAT: generic items			—	201	218	*273*	*306*	238	−127	−136
Explicit bipolar gender self-concept scales										
4. PAQ (bipolar)				—	**723**	**763**	**489**	**483**	**−604**	**−632**
5. Semantic differential (bipolar)					—	**873**	**493**	**545**	**−736**	**−669**
6. BSRI (bipolar)						—	577	502	**−834**	**−773**
Explicit femininity scales										
7. BSRI							—	**687**	−031	−227
8. PAQ femininity								—	−151	−136
Explicit masculinity scales										
9. BSRI masculinity									—	**793**
10. PAQ masculinity										—

Note. Decimal points are omitted. $N = 58$. The bipolar measures (first six in the table) are scored so that higher numbers indicate greater identity as feminine. IAT = Implicit Association Test; PAQ = Personal Attributes Questionnaire; BSRI = Bem Sex Role Inventory. For $N = 58$, *r*s of .259, .364, .443, and .507 are associated, respectively, with two-tailed *p* values of *.05*, .005, **.0005**, and **.00005**.

the ideal American man of previous generations. If these shifts in ideal women and men affect responses to explicit measures, the result would be to observe lower levels of sex differences on Experiment 2's explicit measures than would have been observed in the 1970s.

Indeed, available data do indicate that the mean differences in explicit masculinity and femininity scores observed in the present research are smaller than those reported in the 1970s. Compared with the mid-1970s BSRI and PAQ mean masculinity and femininity scores for college students that are summarized in Lenney's (1991) overview of sex role measures, the mean sex differences reported in Table 3 average only 62% of those from the 1970s. Unfortunately, because implicit measures of gender self-concept were unavailable in the 1970s, there is no way to determine whether sex differences on implicit gender self-concept measures might similarly have shrunk with time. However, if implicit measures are free of societal pressures that might explain changes in self-reported gender self-concepts, then the sex differences observed on present-day implicit measures (Table 2) might not be different from those that would have been observed in the 1970s.

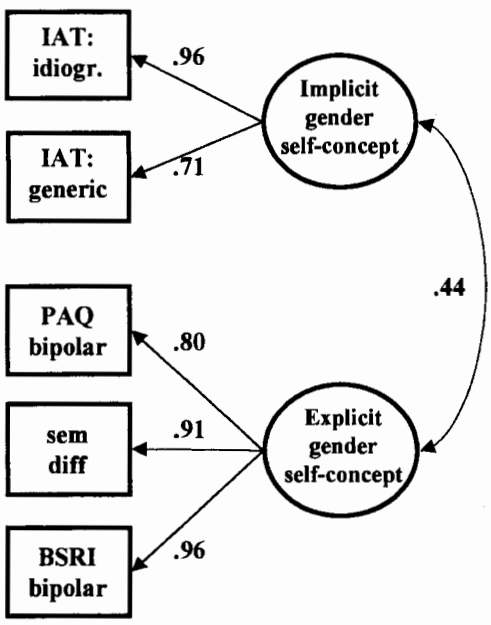

Figure 4. Confirmatory factor analysis of Experiment 2. Applying the RMSEA criterion (see footnote 4), this five-variable, two-factor model had close to fair fit. The fit of a five-variable model that constrained all measures to load on a single-factor was, by comparison, poor. (Data from Experiment 2; $N = 58$.) IAT = Implicit Association Test; PAQ = Personal Attributes Questionnaire; BSRI = Bem Sex Role Inventory; idiogr. = idiographic items; sem diff = semantic differential.

Experiment 2 also supported the results of Experiment 1 by adding to the evidence that implicit and explicit measures assess distinct constructs that are nevertheless positively correlated with one another. The evidence of construct divergence between implicit and explicit measures is, by now, a familiar pattern in studies that include both types of measures (e.g., Bosson et al., 2000; Brauer, Wasel, & Niedenthal, 2000; Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Greenwald & Banaji, 1995; Greenwald et al., 1998; Greenwald et al., 2000).

Experiment 2 additionally showed that the IAT functions similarly as a measure of implicit self-concept in both idiographic and generic formats. The two formats were approximately equivalent in their sensitivity to sex differences. Further, the two formats correlated with each other at a relatively high level ($r = .68$) that approximates the average test–retest reliability of IAT measures in other studies.

Experiment 2's finding of similar results for generic and idiographic formats of self-concept IAT measures suggests that the generic format is likely to be the more efficient. The difference in efficiency is a consequence of the idiographic procedure's requiring an average of perhaps 10 additional minutes to obtain subject-specific information. However, an additional relevant observation is that the idiographic IAT measure tended to correlate more highly with explicit measures than did the generic IAT measure

(see Table 4). The average absolute values of implicit–explicit correlations for the idiographic and generic IAT formats were, respectively, $|r| = .33$ and $|r| = .22$. Related to this, the idiographic measure had a clearly higher weighting on the CFA's implicit self-concept factor (see Figure 4), which indicates that it better defined that latent variable. These observations prompt hesitation in concluding that the two formats should be regarded as equivalent.

Experiment 3: Prediction of Responses to Success and Failure

Experiments 1 and 2 introduced IAT measures of self-esteem and self-concept and provided evidence for their validity in the form of (a) CFAs of implicit (IAT) and self-report measures of the same constructs (Experiment 1 and Experiment 2) and (b) ability to detect known differences in gender self-concepts (Experiment 2). Experiment 3 took a different approach to assessing validity, building on prior findings that self-esteem moderates cognitive reactions to success and failure.

Previous research has shown that low self-esteem persons take negative feedback more to heart than do high self-esteem persons (Brockner, 1983; Brown & Dutton, 1995; Dodgson & Wood, 1998; Greenberg et al., 1992). Compared with persons with high self-esteem, those with low self-esteem are expected to report lower mood and lower self-evaluation of performance after experiencing failure. Experiment 3 tested these expectations by exposing a subset of Experiment 1's subjects to either success or failure after they had completed their measures of implicit and explicit self-esteem.

There is no existing theorization to suggest that implicit and explicit self-esteem should function differently in predicting reactions to success and failure. Therefore, it was expected that Experiment 3 might show the two types of self-esteem to function similarly in predicting reactions to success and failure. At the same time, because implicit and explicit self-esteem appear to be different constructs (present Experiment 1; also Bosson et al., 2000), it was equally plausible that the two types of self-esteem measures would differentially predict reactions to success and failure.

Method

Subjects

Experiment 3's subjects were a subset of 94 (30 men, 64 women) of the subjects who provided usable data for Experiment 1. After providing the measures of Experiment 1, they completed a task that gave them a success or failure experience. Forty-seven subjects were assigned randomly to each of the easy-task (success) and hard-task (failure) conditions.

Procedure

Success–failure variation. After completing the two IATs as described in Experiment 1, subjects completed a paper-pencil task in which they were asked to identify, in a longer list, 20 names that should be familiar because they had appeared in news or entertainment media. To create the experience of success or failure, half of the subjects received a difficult version of the task and half an easy version. Each version contained 60 names, 20 of which had appeared recently in the media. All media names were selected from newspapers and web news summaries. The easy and difficult versions differed in the familiarity of the 20 critical names. Nonmedia names were created through recombinations of the media names. For example, Marilyn Jackson was created as a combination of Marilyn Monroe and Michael Jackson (two names from the easy version of the task).

Dependent measures. After completing the task, subjects were given an answer key and were asked to use it to determine the number of names that they had correctly identified and then to write that number at the bottom of the page. By scoring their own performance, subjects received feedback that was both immediate and anonymous. After receiving feedback for the task, subjects responded to questionnaire items that provided data for four dependent measures, in the following order: Mood—Subjects rated their current mood states using a scale developed by Brown and Dutton (1995). Subjects indicated on a 5-point Likert scale the extent to which their current mood was describable by adjectives such as blue, proud, sad, happy, and worthless. Success—Subjects indicated on a single 7-point scale the extent to which they believed they had succeeded on the just-completed name-identification task. Importance of Task—Subjects indicated on a single 7-point scale how important they thought it was to know current events (a measure of the importance of the ability measured by their just-completed task). Level of Aspiration for Future Performance—Subjects were informed that the task would be repeated with different names. They were then asked to indicate how many of the 20 critical names on the upcoming task they would hope to identify correctly.

Implicit and Explicit Self-Esteem Measures

The implicit self-esteem measure was an equally weighted averaged of Experiment 1's two IATs, computed by standardizing each measure prior to averaging the two. The explicit self-esteem measure was a similarly standardized composite of Experiment 1's RSES and thermometer measures. These two explicit measures were selected from the four used in Experiment 1 because they best represented the explicit self-esteem factor of Experiment 1's CFA (see Figure 3).

Results and Discussion

The manipulation of task difficulty succeeded in producing the desired variations in actual and perceived success at the name-identification task. Subjects had on average 15.4 of 20 correct responses in the easy condition, compared with 6.4 of 20 in the difficult condition, $t(92) = 11.64$, $p = 10^{-19}$. Even more importantly, subjects reported feeling much more successful after completing the easy ($M = 5.9$) than the difficult version ($M = 3.2$), $t(83) = 7.48$, $p = 10^{-10}$. If any of the self-esteem measures had been correlated with performance at the name-identification task, the success–failure manipulation would have been compromised. Fortunately, performance (number correct) was uncorrelated with implicit or explicit self-esteem within either the success (easy) condition (rs = .14 and −.04, respectively) or the failure (difficult) condition (rs = −.12 and −.04).

If high self-esteem provides cognitive protection against the effects of failure feedback, then, compared with subjects with low self-esteem, subjects with high self-esteem should show smaller effects of the success–failure manipulation on the four measures of its impact: (a) judgment of having failed or succeeded, (b) current mood, (c) judged importance of the ability measured by the task, and (d) expected future performance at the task. This prediction calls for an interaction effect of self-esteem and success–failure on the four measures, such that higher levels of self-esteem should be associated with smaller differences between success and failure conditions on each measure. This interaction-effect prediction was tested with a two-step hierarchical regression analysis for each of the four dependent measures. Self-esteem and task feedback (suc-

cess or failure) were entered on the first step of the analysis to estimate their main effects. On the second step, the interaction effect was tested by entering as a predictor the multiplicative product of self-esteem and success–failure (the latter dummy-coded as 0 or 1). Results of the analyses of the four measures are graphed in Figure 5.

Effects of Task Feedback (Success–Failure)

Effectiveness of the success–failure manipulation was indicated by the occurrence of expected effects of the manipulation on all four measures. The largest effect, not surprisingly, was the already described effect on judgment of success at the task (Cohen's **d** = 1.60, $p = 10^{-10}$). The success condition also produced higher means on the other three measures: posttask mood (**d** = .48, p = .02), importance of the ability assessed by the name identification task (**d** = .54, p = .01), and performance aspiration for a repetition of the task (**d** = .51, p = .02). The theoretical significance of these main effects of the success–failure manipulation is that they establish the conditions needed to assess the interaction-effect prediction. That is, for any measure that shows a main effect of success–failure, the self-esteem-buffering hypothesis predicts greater difference between success and failure conditions for participants with low self-esteem than for those with high self-esteem.

Main Effects of Measured Self-Esteem

Implicit self-esteem had no main effects, and only one main effect of explicit self-esteem was observed, an effect of explicit self-esteem on posttask mood (Figure 5, second panel on left). Regardless of task feedback condition, subjects high on the explicit self-esteem measure had more positive posttask moods, and this was a strong effect, regression $\beta = .51$, $p = 10^{-7}$. However, it is plausible that this result indicates only that the mood measure (self-ratings of positive feelings) and the explicit self-esteem measures (self-ratings of other positive attributes) called for similar types of self-positivity judgments.

Interaction of Self-Esteem and Success–Failure

The focus of theoretical interest in Experiment 3 was the analysis of interaction effects involving success–failure and self-esteem. If high self-esteem participants have a cognitive protection against negative feedback, then (relative to subjects with low self-esteem) they should show reduced differences between success and failure conditions on measures that were affected by success versus failure. (This includes all four of Experiment 3's rating measures that were collected following task feedback.) For the data sets plotted in Figure 5, these interactions should appear as a pair of slopes that define a > pattern (converging to the right). Figure 5 has three interaction effects that show this pattern strongly enough to warrant notice. In the upper left panel of Figure 5, the difference between success and failure conditions in rated success was smaller for high- than for low-explicit-self-esteem subjects, interaction $F(1, 80) = 2.60$, partial $r = .18$, p = .11. In the two lower right panels, the >-shaped pattern can be seen for the measures of task importance, interaction $F(1, 90) = 3.84$, partial $r = .20$, $p = .05$, and future aspiration, interaction $F(1, 90) = 3.64$, partial $r = .20$, $p = .06$.

In summary, Experiment 3's task feedback manipulation succeeded in establishing distinct experiences of success and failure. For implicit self-esteem, the expected effect of high self-esteem in buffering effects of failure was observed for two of the four posttask rating measures. Although p values for these two effects straddled the $p = .05$ level that is often treated as a boundary between noteworthy and ignorable results, any inclination to dismiss these findings should be tempered by noting that these two effects agreed with prediction in both direction and shape. To elaborate: A significant interaction effect of the type tested in Experiment 3 could have been produced by either a < or > pattern of the two regression slopes. The occurrence of the predicted slope directions might therefore justify halving the computed p values. Further, even with occurrence of the predicted directions of slopes, it would have been possible for these significant interaction effects to occur with slopes for the failure condition elevated above those for the success condition, rather than in the predicted pattern of success slope elevated relative to failure slope. Therefore, the finding that predicted interaction effects occurred with the predicted direction and shape prompts more confidence than the stated p values might otherwise appear to warrant.

Additional analyses were conducted using individual self-report and IAT measures in place of the standardized composite predictor measures that are presented in Figure 5. With one exception, the results were entirely consistent with the patterns shown in Figure 5. The exception was that the analysis of the mood dependent variable with the thermometer measure of explicit self-esteem produced an interaction effect ($p = .06$) that was opposite in direction from prediction.

Even though the four analyses for explicit self-esteem yielded no findings for which the p value dropped below .05, it can be seen on the left side of Figure 5 that three of the four explicit measures—all except the mood measure—did display the predicted > interaction shape. The appearance of statistically stronger effects in the implicit measures than in the explicit self-esteem measures remains, for the present, unexplained. There was no theoretical reason, a priori, to expect that implicit and explicit self-esteem should produce different patterns or magnitudes of the predicted effect of buffering against failure.

Additional Data: Test–Retest Reliability

Experiment 1 showed that two self-esteem IATs with different sets of positive and negative items (affective words and evaluative trait words) correlated positively with each other ($r = .43$), and Experiment 2 showed that two gender self-concept IATs with different types of self and other items (generic pronouns vs. idiographic subject-generated items) correlated with each other ($r = .68$). These two correlations provide a measure of stability of IAT self-concept measures in the form of correlations between parallel measures administered during the same session. This section describes additional data that provided a test–retest reliability estimate based on the generic form of the self-esteem IAT.

Subjects were students at University of Washington who participated in two of three experiments in exchange for course credit. All three experiments included a generic self-esteem IAT and the RSES. These additional experiments, which had the aim of examining relations of self-esteem to minimal group effects (Farnham, 1999), were run in an academic term after the completion of

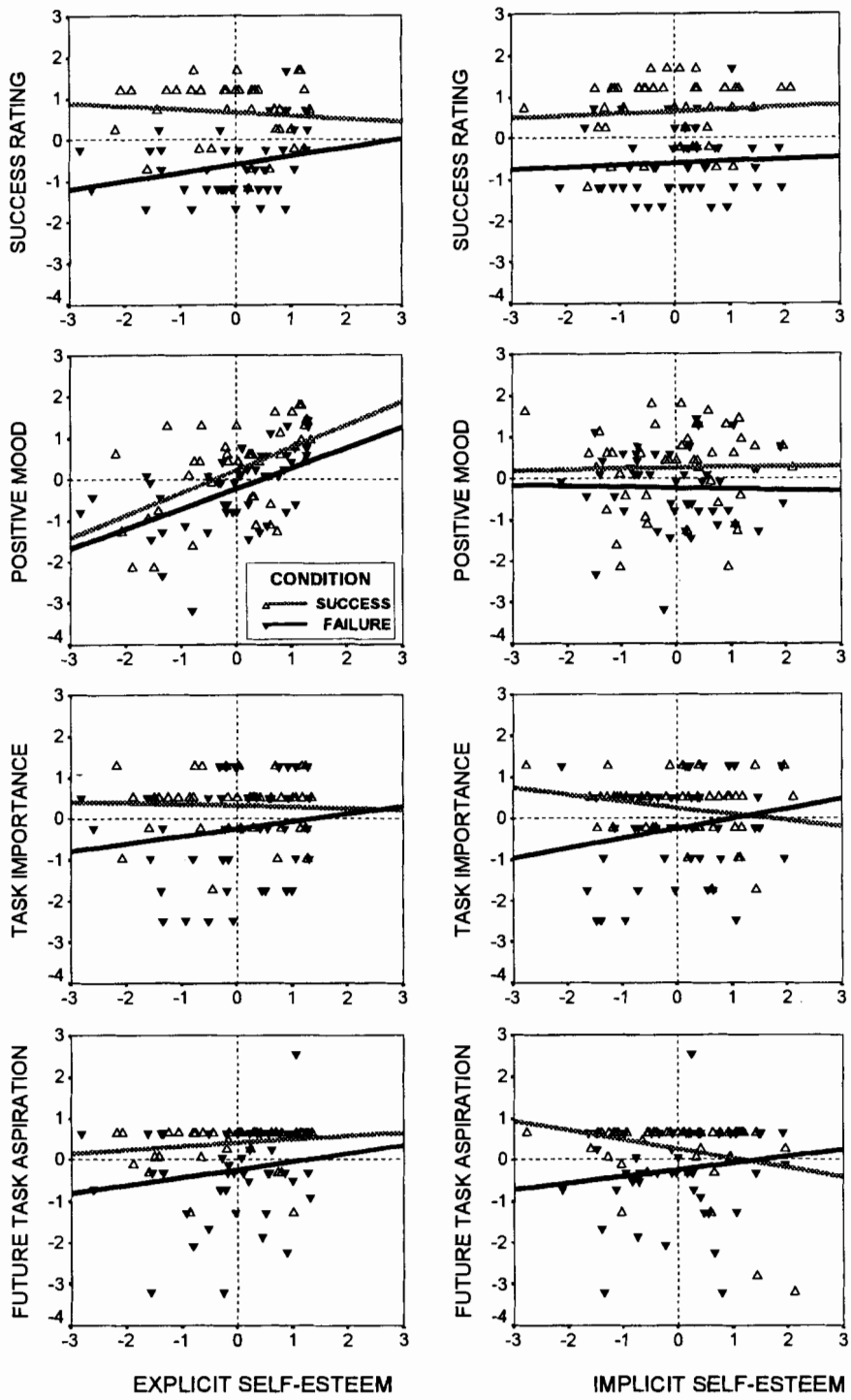

Figure 5. Regression analyses of four dependent measures as a function task condition (success vs. failure) and measured self-esteem (explicit self-esteem in left panels, implicit self-esteem in right panels). Separate regression slopes are plotted for each condition. Interaction effects have the predicted > shape at statistically noticeable levels in the upper left panel and the lower two right panels. All measures are standardized to facilitate comparisons among the analyses.

Experiments 1–3. Fourteen subjects were excluded from the test–retest reliability analysis for having error rates in excess of 20% for a critical block in at least one of their two IAT measures. Seven additional subjects were excluded for having less than 10 years of experience speaking English, and 1 subject was excluded as an extreme outlier (standardized residual of 3.3 for the regression of second IAT on first). Data were analyzed for the remaining 44 subjects (36 women and 8 men; these included 20 Asians, 22 Caucasians, and 2 members of other racial groups).

Subjects completed a generic form of the self-esteem IAT and the RSES twice during the same 10-week academic term. The generic self-esteem IAT used sets of six pronouns to define self and other and used sets of eight words to define pleasant and unpleasant (see Appendix for the full sets of items). The average time between sessions was 8 days (SD = 14.9), and 50% of the subjects did both IATs on the same day (i.e., they participated in two different experiments on the same day). Order effects were tested separately for both the first and second IATs using a 2 (order of critical blocks: self+positive first or second, between-subjects) × 2 (type of critical block: self+positive vs. self+negative, within-subjects) ANOVA. Both IATs showed an order effect of the sort previously observed (e.g., Greenwald et al., 1998, Experiment 1) such that larger IAT effects were observed when the first critical block presented the task combination that was easier for almost all subjects (i.e., self+pleasant). This order effect was significant for the first IAT taken, $F(1, 42) = 10.12$, $p = .003$, but not for the second IAT, $F(1, 42) = 2.06$, $p = .16$.

Reliability

Test–retest reliability of the RSES was $r = .65$ ($N = 44$, $p = 10^{-6}$). Test–retest reliability of the IAT self-esteem measure was $r = .52$, ($N = 44$, $p = .0003$).

Other Analyses

Consistent with previous observations of small correlations between implicit and explicit self-esteem measures (e.g., present Experiment 1; Bosson et al., 2000), the IAT and the RSES were correlated weakly at both testings—$r = .12$, $p = .46$ at Time 1; $r = .21$, $p = .17$ at Time 2. For both the IAT and RSES, self-esteem scores were slightly lower for the second than the first test, but these differences were not statistically significant. The interval between testings did not affect difference between the two IAT measures ($r = -.21$, $p = .17$). However, length of time between two administrations was correlated with the drop in score from first to second taking for the RSES ($r = -.42$, $p = .005$). That is, the more separated the two measures, the lower were the explicit self-esteem scores on the second test. Although potentially interesting, this result also has some mundane interpretations—for example, explicit self-esteem may have declined as final examinations neared. Sample sizes were too small to provide adequate tests of race or sex effects that could be compared with those of Experiment 1.

General Discussion

Summary of Findings

The interpretation of self-report measures—of self-esteem, self-concept, or other constructs—is potentially complex because such measures can intermix, in unknown proportions, both valid indication of self-concept and self-presentational distortions. The IAT is an indirect measure that does not rely on introspection and has been found to be low in susceptibility to self-presentational distortion (Kim & Greenwald, 2000). Although these properties make the IAT a potentially desirable measure for use in research, its establishment as a useful research measure depends on its meeting the usual psychometric standards for individual difference measures. The present studies provided such evidence—concerning psychometric criteria of stability and three forms of construct validity.

The present three experiments provide evidence of IAT measures' stability in the form of substantial correlations between alternate IAT test forms in Experiment 1 and Experiment 2 and a test–retest reliability estimate from the additional data collection. The average of these three stability estimates (average $r = .55$, using r to Z conversion) is slightly low for measures that are intended for use as accurate individual-respondent assessments. Nevertheless, this stability is satisfactorily high for research uses in correlational studies. The average r of .55 is also close to the average of stability estimates of IAT attitude measures that were previously reported by Greenwald et al. (1998) and by Dasgupta and Greenwald (2000; average $r = .64$; summarized in the introduction of this article).

The present studies provided construct validity evidence in three forms: (a) known groups validity, (b) predictive validity, and (c) discriminant validity. Known groups validity evidence was provided by Experiment 2's demonstration that IAT measures were highly sensitive to known differences between men and women in masculinity and femininity of self-concept. Predictive validity was shown in Experiment 3's finding that IAT-measured implicit self-esteem predicted an expected buffering (for those high in self-esteem) on two of four measures of cognitive reactions to manipulated success versus failure. Discriminant validity appeared in findings from Experiments 1 and 2 and from the additional data collection—in all of these, low correlations between IAT measures and explicit measures indicated that IAT measures of implicit self-esteem and self-concept measure something different from what is measured by explicit (self-report) measures of self-esteem and self-concept.

Experiment 1 produced the unexpected observation that the correlation between implicit and explicit self-esteem was higher for Caucasian men than for Asian and/or female subjects. Because this result was obtained from exploratory analyses, it was not possible to give it a confidently interpretable p value. Nevertheless, this result seems worthy of eventual follow-up because of more general interest in understanding factors that moderate the agreement between implicit and explicit measures.

In summary, the present experiments provided initial evidence that IAT measures of implicit self-esteem and implicit self-concept (a) have psychometric properties of stability and validity that justify their use in research settings and (b) define constructs that are distinct from, although correlated with, nominally the same constructs as measured by self-report.

Other Research Using IAT Measures of Self-Esteem or Self-Concept

The first of the present studies was conducted in fall, 1996, and was reported informally well before it was submitted for publication as part of the present article. Consequently, implicit self-esteem and self-concept IAT measures have been used in other studies not only by the present authors, but also by several others. Farnham et al. (1999) and Bosson et al. (2000) have reported additional studies that compared IAT and explicit measures of self-esteem. The Farnham et al. (1999) report briefly summarized studies that used a single IAT measure of self-esteem, along with six self-report measures of self-esteem. Like the present research, Farnham et al. (1999) reported weak positive correlations between the IAT and explicit self-esteem measures (average r = .18). Bosson et al. (2000) reported the most complete data set yet obtained that included both measures of implicit self-esteem (seven measures) and explicit self-esteem (four measures). Perhaps disappointingly, the seven measures of implicit self-esteem were not positively correlated with one another. Nevertheless, the Bosson et al. (2000) research did provide evidence for the value of the IAT as a measure of implicit self-esteem. In particular, they found (a) that the IAT had the highest test–retest reliability (r = .69) of the seven implicit self-esteem measures that they investigated, (b) that, similar to the present results, the IAT was consistently and weakly correlated with explicit self-esteem measures (average r = .21), and (c) that the IAT was positively and significantly correlated with three of six behavioral criterion measures for which Bosson et al. had described an a priori expectation of correlation with measures of self-esteem.

In a doctoral dissertation, Farnham (1999) reported the first use of the IAT as an experimental dependent measure. Each subject was asked to imagine being a member of one of two fictitious four-person groups. An IAT measure of attitude toward the group used the sets of four first names to represent each group and used pleasant and unpleasant words to represent the attributes of positive and negative valence. Similarly, an IAT measure of group identification used the two sets of first names along with pronouns representing the concepts of self and other. The IAT identification measure showed that the fictitious membership group had implicitly become an ingroup (self was associated more with the membership group than with the other group); the IAT attitude measure showed the expected ingroup favoritism. In another dissertation, Haines (1999) similarly used an IAT self-concept measure as an experimental dependent measure. Haines's experiment found that women assigned to a powerful role in a simulation game increased their association of self with the attribute of dominance.

Swanson, Rudman, and Greenwald (in press) used IAT attitude and self-concept measures in correlational investigations of groups that were defined by either an addictive habitual behavior (smoking) or a nonaddictive habitual behavior (vegetarianism). IAT measures of association of self with cigarettes or with nonmeat protein were correlated significantly with the respective behavioral measures of smoking and diet and were correlated more strongly with those self-report behaviors than were IAT attitude measures.

Gender identity has been the subject of several investigations that have used IAT measures of self-concept or self-esteem. Rudman et al. (2000) investigated relations among self-esteem, gender identity, and gender stereotypes, finding that both women and men have own-gender-favorable implicit stereotypes that can be interpreted in terms of associations of self both with own gender and with favorable traits. Rudman and Heppen (2000) reported relations between women's romantic identity and self-esteem (both measured with the IAT). Their results led them to conclude that associating self with romance can be a hindrance to women's career progress. Nosek, Banaji, and Greenwald (2000) found that both men and women associated math more with male than with female gender and that women showed less association of self with mathematics than did men. Cook, Park, and Greenwald (2000) extended the Nosek et al. (2000) result, reporting that men also associated self more with science and engineering fields than did women.

In the first study that used a self-concept IAT in a clinical setting, Gemar, Segal, Sagrati, and Kennedy (2000) administered a mood manipulation to (a) formerly depressed patients, (b) currently depressed patients, and (c) never-depressed controls. In their no-mood-manipulation control condition, a self-esteem IAT measure revealed higher self-esteem for formerly depressed than for currently depressed groups. However, in the sad-mood (experimental treatment) condition, IAT scores of formerly depressed patients were indistinguishable from those of currently depressed patients. Gemar et al. (2000) concluded that the IAT was sensitive to former depressives' susceptibility to automatic negative effects of mood variations.

A recent development in investigations with IAT measures is the use of *balanced identity* designs (Greenwald et al., 2000). Balanced identity designs test strengths of associations within all pairs of three concepts, one of which is self. Greenwald et al. (2000) described a set of correlational tests that assess the extent to which measures of association strength within triads of concepts are cognitively consistent, or balanced. In the first study to use a balanced identity design, Farnham and Greenwald (1999) examined the associations among self (vs. other), female (vs. male), and pleasant (vs. unpleasant) for women college students. In a theoretical article, Greenwald et al. (2000) summarized this and four additional studies that used balanced identity designs (Banaji, Nosek, Greenwald, & Rosier, 2000; Mellott & Greenwald, 1999; Nosek et al., 2000; Rudman et al., in press). In all five studies, data that were obtained with IAT measures of association strength conformed to the pattern expected for cognitively consistent triads, whereas data obtained with explicit measures of association strength did not display such consistency. These findings from balanced identity designs add to the case for construct divergence between constructs measured by the IAT and those measured by self-report.

Conclusion

In 1995, Greenwald and Banaji concluded that measurement of implicit constructs had "not yet been achieved in the efficient form needed to make research investigation of individual differences in implicit cognition a routine undertaking." Optimistically, they forecast that "When such measures do become available, there should follow the rapid development of a new industry of research on implicit cognitive aspects of personality" (1995, p. 20). The realization of that forecast no longer seems so distant as it did in 1995.

References

Aidman, E. V. (1999). Measuring individual differences in implicit self-concept: Initial validation of the Self-Apperception Test. *Personality and Individual Differences, 27,* 211–228.

Banaji, M. R., Nosek, B. A., Greenwald, A. G., & Rosier, M. (2000). *Implicit and Explicit Self-Esteem and Group Membership.* Manuscript in preparation.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893–912.

Bargh, J. A., & Tota, M. E. (1988). Context-dependent automatic processing in depression: Accessibility of negative constructs with regard to self but not others. *Journal of Personality and Social Psychology, 54,* 925–939.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, and Computers, 18,* 299–303.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162.

Blair, I., & Ma, J. (1999). *Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery.* Manuscript submitted for publication.

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.

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

Brockner, J. (1983). Low self-esteem and behavioral plasticity. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 4, pp. 237–271). Beverly Hills, CA: Sage.

Brown, J. D., & Dutton, K. A. (1995). The thrill of victory, the complexity of defeat: Self-esteem and people's emotional reactions to success and failure. *Journal of Personality and Social Psychology, 68,* 712–722.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cook, K., Park, L. E, & Greenwald, A. G. (2000, June). *Implicit associations and women's commitment to math, science and engineering.* Paper presented at meetings of the American Psychological Society, Miami Beach, FL.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (in press). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science.*

Dasgupta, N., & Greenwald, A. G. (2000). *Exposure to admired group members reduces automatic intergroup bias.* Manuscript submitted for publication.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36,* 316–328.

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

Dodgson, P. G., & Wood, J. V. (1998). Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *Journal of Personality and Social Psychology, 75,* 178–197.

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

Farnham, S. D. (1998). FIAT for Windows [Computer software]. Seattle, WA: Author. Available: http://www.hive-mind.com/shelly/IAT/ [1998, June 2].

Farnham, S. D. (1999). *From implicit self-esteem to in-group favoritism.* Unpublished doctoral dissertation, University of Washington, Seattle, WA.

Farnham, S. D., & Greenwald, A. G. (1999, June). *In-group favoritism = implicit self-esteem × in-group identification.* Paper presented at meetings of the American Psychological Society, Denver, CO.

Farnham, S. D., Greenwald, A. G., & Banaji, M. R. (1999). Implicit self-esteem. In D. Abrams & M. Hogg (Eds.), *Social identity and social cognition* (pp. 230–248). Cambridge, MA: Blackwell Publishers.

Fazio, R. H., Powell, M. C., & Herr, P. M. (1983). Toward a process model of the attitude–behavior relation: Accessing one's attitude upon mere observation of the attitude object. *Journal of Personality and Social Psychology, 44,* 723–735.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2000). *Contributions of effortful and automatic measures of cognition to a risk marker for depressive relapse/recurrence: The Implicit Association Test in depression.* Manuscript submitted for publication.

Greenberg, J., Solomon, S., Pyszczynski, T., Rosenblatt, A., Burling, J., Lyon, D., Simon, L., & Pinel, E. (1992). Why do people need self-esteem? Converging evidence that self-esteem serves an anxiety-buffering function. *Journal of Personality and Social Psychology, 63,* 913–922.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35,* 603–618.

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A. & Mellott, D. S. (2000). *A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept.* Manuscript submitted for publication.

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., & Pratkanis, A. R. (1984). The self. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 129–178). Hillsdale, NJ: Erlbaum.

Gur, R. C., & Sackeim, H. A., (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37,* 147–169.

Haines, E. L. (1999). *Elements of a social power schema: Gender standpoint, self-concept, and experience.* Unpublished doctoral dissertation, City University of New York.

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology, 35,* 512–559.

Kihlstrom, J. F., & Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp 1–47). Orlando, FL: Academic Press, Inc.

Kim, D-Y., & Greenwald, A. G. (2000). *Voluntary controllability of implicit cognition: Can an implicit measure (the IAT) of attitudes be faked?* Manuscript submitted for publication.

Kitayama, S., & Karawawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin, 23,* 736–742.

Lenney, E. (1991). Sex roles: The measurement of masculinity, femininity, and androgyny. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 573–660). San Diego, CA: Academic Press.

Lowery, B., & Hardin, C. D. (1999, June). *Social tuning effects on automatic racial prejudice.* Paper presented at the annual meeting of the American Psychological Society, Denver, CO.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power

analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63–78.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive.* New York: Appleton-Century-Crofts.

Mellott, D. S., & Greenwald, A. G. (1999). *But I don't feel old! Implicit self-esteem, age identity and ageism in the elderly.* Unpublished manuscript, University of Washington, Seattle, WA.

Mellott, D. S., & Greenwald, A. G. (2000, May). *Measuring implicit ageism: Do the Implicit Association Test and semantic priming measure the same construct?* Paper presented at meetings of the Midwestern Psychological Association, Chicago, IL.

Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.

Nosek, B., & Banaji, M. R. (2000). *Measuring implicit social cognition: The single category association task.* Unpublished manuscript, Yale University, New Haven, CT.

Nosek, B., Banaji, M. R., & Greenwald, A. G. (2000). *Math = Male, Me = Female, therefore Math ≠ Me.* Unpublished manuscript, Yale University, New Haven, CT.

Nuttin, J. R. (1985). Narcissism beyond Gestalt awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353–361.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (in press). Implicit attitudes and racism: The effect of word familiarity and frequency in the Implicit Association Test. *Social Cognition.*

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29,* 1049–1071.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes. Measures of social psychological attitudes* (Vol. 1., pp. 17–59). San Diego, CA: Academic Press.

Pelham, B. W., & Hetts, John, J. (1999). *Implicit self-evaluation.* Unpublished manuscript.

Pelham, B. W., & Swann, W. B. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology, 57,* 672–680.

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology, 59,* 475–486.

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

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35,* 677–688.

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

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). New York: Academic Press.

Rudman, L. A., Ashmore, R. D., & Gary, M. (1999). *Implicit and explicit prejudice and stereotypes: A continuum model of intergroup orientation assessment.* Manuscript submitted for publication.

Rudman, L. A., & Glick, P. (1999). *Prescriptive gender stereotypes and backlash toward agentic women.* Manuscript submitted for publication.

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

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

Rudman, L. A., & Heppen, J. (2000). *Someday my prince will come: Implicit romantic fantasies and women's avoidance of power.* Manuscript submitted for publication.

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

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.

Spence, J. T., & Helmreich, R. L. (1979). The many faces of androgyny: A reply to Locksley and Colten. *Journal of Personality and Social Psychology, 37,* 1032–1046.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-femininity. *Journal Supplement Abstract Service Catalog of Selected Documents in Psychology, 4,* 43–44.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (in press). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion.*

Taylor, S. E., & Brown, J. D. (1984). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.

Tedeschi, J. T., Schlenker, B. R., & Bonoma, T. V. (1971). Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist, 26,* 685–695.

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

Wylie, R. (1974). *The self-concept: A review of methodological considerations and measuring instruments, Volume 1.* Lincoln: University of Nebraska Press.

(Appendix follows)

Appendix

Items Used in the IATs for All Experiments

Items for Experiments 1 and 3

Affective		Evaluative		Idiographic (Me or Not-me)	
Positive	Negative	Positive	Negative	Items	Examples
caress	abuse	smart	stupid	birth day	Feb 19
cuddle	agony	bright	ugly	birth year	1963
diamond	assault	success	failure	city 1	London
glory	brutal	splendid	awful	city 2	Boston
gold	corpse	valued	useless	country	Italy
health	death	noble	vile	first name	Jennifer
joy	filth	strong	weak	gender	female
kindness	killer	proud	ashamed	ethnicity 1	Chinese
lucky	poison	loved	hated	ethnicity 2	Irish
peace	slum	honest	guilty	handedness	left-handed
sunrise	stink	competent	awkward	last name	Carter
truth	torture	worthy	rotten	middle name	Donald
warmth	vomit	nice	despised	state	Maine
				religion	Hindu
				phone number	nnn-nnnn
				street name	Oak St
				Social Security no.	nnn-nn-nnnn
				zip code	98105

Items for Experiment 2

Idiographic items	Generic items (pronouns)		Gender self-concept items	
	Self	Other	Feminine	Masculine
first name	I	they	gentle	competitive
middle name	me	them	warm	independent
last name	my	their	tender	forceful
city	mine	it	sensitive	strong
state	self	other	sympathetic	confident
country			soft	aggressive

Items for Additional Test–Retest Reliability Study

Self	Other	Positive	Negative
myself	other	rainbow	pain
mine	them	happy	death
me	their	smile	poison
my	they	joy	grief
myself	them	warmth	agony
self	other	pleasure	sickness
		paradise	tragedy
		sunshine	vomit

Note. The Self and Other categories show some items listed twice because these items appeared twice as often as items listed only once.

Received March 28, 2000
Revision received July 11, 2000
Accepted July 13, 2000 ■