Social Cognition, Vol. 17, No. 4, 1999, pp. 437–465

MEASURING THE AUTOMATIC COMPONENTS OF PREJUDICE: FLEXIBILITY AND GENERALITY OF THE IMPLICIT ASSOCIATION TEST

LAURIE A. RUDMAN
Rutgers University

ANTHONY G. GREENWALD, DEBORAH S. MELLOTT,
AND JORDAN L. K. SCHWARTZ
University of Washington

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) provides a flexible measure of the automatic associations underlying implicit prejudice. Results of three experiments showed strong evidence of implicit prejudices based on religious ethnicity (Jewish vs. Christian), age (young vs. old), and nationality (American vs. Soviet). Subjects responded more rapidly to tasks that obliged association of ingroup tokens to pleasant attributes and outgroup tokens to unpleasant attributes than to ones that obliged the complementary associations. In addition, the findings of three experiments were consistent with the hypothesis that IAT effects are independent of self-reported stimulus familiarity differences. These results support the construct validity and the generality of the IAT method in implicit prejudice research.

Despite the fundamentally egalitarian tenets on which democratic societies are based, the tendency to form attitudes that favor ingroup members and disfavor outgroup members is pervasive (Brewer, 1979; Selznick & Steinberg, 1969). The differential evaluation of ingroup versus outgroup members is the basis of prejudice, one definition of which is a "negative attitude toward a person or group based upon a social comparison process in which the individual's own group is taken as a positive point of reference" (Jones, 1972, p. 3). Social-cognitive explanations for the tenacity of prejudice stem from the assumption that catego-

This research was partially supported by National Research Service Award 1 F32 MH11632-01 from National Institute of Mental Health to the first author, and Grant SBR-9422242 from National Science Foundation and Grant MH 41328 from National Institute of Mental Health to Anthony G. Greenwald.

Address correspondence to Laurie A. Rudman, Department of Psychology, Tillett Hall, Rutgers University, Piscataway, NJ, 08854-8040; E-mail: rudman@rci.rutgers.edu.

rization based on group differences (e.g., age, race, gender) is an automatic and universal process which inevitably leads to differential evaluation of group members (i.e., prejudice; Duckitt, 1992; Perdue, Dovidio, Gurtman, & Tyler, 1990). When this differential evaluation is unconscious, we will refer to it as *implicit* prejudice (Wittenbrink, Judd, & Park, 1997). It is based on the automatic association between group members and negative evaluation. As such, it is conceptually distinct from explicit prejudice (i.e., affect that is consciously antipathetic; Allport, 1954). The hallmark of implicit prejudice is that it operates without individuals' conscious awareness. It may therefore unintentionally influence judgments and behaviors in ways detrimental to members of stigmatized groups (Greenwald & Banaji, 1995).

Intriguingly, at the same time that social-cognitive research is recording the universality of implicit prejudice (e.g., Devine, 1989; see Greenwald & Banaji, 1995, for a review), survey research is documenting dramatic decreases in explicit prejudice and stereotypes (e.g., Judd, Park, Ryan, Brauer, & Kraus, 1995; Spence & Hahn, 1997). However, the reliance on self-report measures in prejudice research provokes a question as to whether the research accurately assesses attitudes or self-presentation. Self-reports provide the opportunity to project distorted expressions that may present a favorable impression (Dovidio & Fazio, 1992). They also presume that respondents have access to their attitudes, including those that are complex, ambivalent, or unconscious. The widespread use of self-report instruments implies that researchers believe respondents to be both willing and able to report their attitudes on demand.

Because prejudices are controversial attitudes, their public expression has implications both for self-regard and evaluation by others. Efforts to augment the validity of self-reports have taken both methodological and conceptual approaches. The methodological approach includes the bogus pipeline, which attempts to persuade respondents that experimenters can detect deception (Jones & Sigall, 1971), and the use of context effects embedded in attitude measures, which obscures the focal attitude object (e.g., Sniderman & Carmines, 1997). The conceptual approach is exemplified by the Modern Racism Scale (MRS; McConahay, 1986). Because changes in public norms dictate that "old-fashioned" prejudice is to be shunned, the MRS measures racism based on "democratic values" rather than blatant hostility. Although these methods may encourage honesty, they do not prevent response editing. Indeed, the MRS has been shown to be reactive in college student samples (Fazio, Jackson, Dunton, & Williams, 1995). Moreover, to the extent that individuals are motivated to maintain a positive, nonprejudiced image to themselves (Greenwald & Breckler, 1985), attempts to induce respondents to "confess their prejudices" may not lead to accurate responses on direct measures.

JUDGMENT LATENCY MEASURES: AN INDIRECT ALTERNATIVE

Indirect methods of measuring prejudice are available. The most promising of these assess judgment latencies for tasks designed to be facilitated (or inhibited) by respondents' attitudes. Attitude-consistent judgments are performed faster than attitude-inconsistent judgments because they are relatively automatic and effortless. Unlike self-report measures, these methods do not depend on the assumption that respondents are willing and able to report their attitudes. Instead, the automatic activation of an attitude toward a social object facilitates or interferes with a subsequent judgment task, influencing the speed and accuracy of decision making (Fazio, 1995). This approach allows for the assessment of implicit prejudice and stereotypes—constructs that respondents are unaware of as an influence over judgments and behavior (Greenwald & Banaji, 1995). Although research in implicit attitudes and beliefs is relatively recent, it appears to be consistent with the social-cognitive framework. Despite what subjects report on questionnaire measures, a strong tendency to automatically distinguish and differentially evaluate ingroup versus outgroup members emerges when response latency tools are used (e.g., Fazio et al., 1995; Perdue et al., 1990). Notably, these attitudes have been shown to have negative implications for intergroup behavior, particularly when actions are spontaneous or uncontrolled (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995).

THE GENERALITY AND FLEXIBILITY OF THE IAT

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a recent addition to the arsenal of response latency tools. Greenwald et al. (1998) suggested that this method could indirectly assess a wide range of implicit attitudes. In Greenwald et al.'s (1998) original paper, the IAT revealed evidence of implicit racism on the part of White Americans toward African Americans, irrespective of subjects' scores on explicit measures. It also showed evidence of known groups validity when the intergroup attitudes of Koreans and Japanese subjects were IAT-assessed. Additional research has shown the IAT to be an effective measure of implicit gender beliefs (e.g., Rudman, Greenwald, & McGhee, 1998; Rudman & Kilianski, in press) and self-esteem (Farnham, Banaji, & Greenwald, 1999). The main objective of the present research was to extend the IAT's usefulness as a general measure of implicit prejudice and stereotypes.

The IAT requires individuals to respond to four types of words, using only two response keys. In prejudice or stereotyping research, subjects

categorize ingroup and outgroup tokens (*target concepts*), along with stimuli representing the poles of an *attribute dimension*. To assess implicit *prejudice*, the attribute dimension is evaluative and consists of pleasant versus unpleasant words (Greenwald et al., 1998). When group tokens and valenced words are evaluated similarly, mapping them onto the same response key is considerably easier than when group tokens and valenced words are evaluated differently. This speed or latency difference (*IAT effect*) measures the extent to which positive and negative evaluation are automatically associated with ingroups and outgroups, respectively. To assess implicit *stereotypes*, the attribute dimension consists of stereotypic and nonstereotypic words (e.g., Rudman et al., 1998). When group tokens and attributes are strongly associated in memory, mapping them onto the same response key is considerably easier than when tokens and attributes are weakly associated.

By varying the ingroup and outgroup identities, the IAT is easily configured to assess a wide variety of implicit attitudes and stereotypes. This flexibility allows researchers to go beyond self-report instruments to assess unconscious attitudes toward (and beliefs about) any social group that can be represented uniquely by a small set of identifying stimuli.

THE INFLUENCE OF PRIOR EXPOSURE ON IAT EFFECTS

Because ingroup tokens may be naturally more familiar than outgroup target tokens, a difference in stimulus familiarity could jeopardize the IAT's construct validity in prejudice research. For example, Greenwald et al. (1998, Experiment 3) found evidence, using the IAT, of implicit prejudice toward African Americans on the part of White subjects. The stimuli used to represent Whites (e.g., Greg, Sara) may have been more familiar to subjects than the stimuli used to represent African Americans (e.g., Malik, Latisha). Because familiar stimuli tend to be preferred over unfamiliar stimuli (Zajonc, 1968), IAT effects could be spuriously inflated by differences in prior exposure to stimuli. It is important to note that the IAT also showed strong evidence of implicit attitudes using nonsocial stimuli that were ostensibly equally familiar (e.g., flowers vs. insects; Greenwald et al. 1998, Experiment 1). Nonetheless, an important aim of the present research was to specifically test whether stimulus familiarity influences IAT effects when stimuli are social.

OVERVIEW OF THE RESEARCH

In seeking to establish the broad usefulness of the IAT for measuring prejudice, the present research applied the IAT's measurement strategy

to three ingroup-outgroup contrasts: Jewish-Christian in Experiment 1, old-young in Experiment 2, and Soviet-American in Experiment 3. The expectation in each case was that implicit ingroup favoritism (and its corollary, implicit outgroup derogation) would be revealed by the IAT, even though it might not be detected on self-report measures. In each study, steps were taken to establish that the IAT effect merited interpretation as implicit prejudice, rather than being an artifact of greater familiarity with ingroup exemplars (Greenwald et al., 1998; Zajonc, 1968). The strategy for examining prior exposure effects on the IAT was three-fold. Experiment 1 used ingroup and outgroup tokens (Jewish and Christian surnames) that were equally frequent in the language. Experiment 2 directly assessed the familiarity of young and old names and statistically controlled for this potential confound in analyses of implicit ageism and age stereotyping. Experiment 3 manipulated the familiarity of tokens to directly test the effect of prior exposure on implicit prejudice based on nationality. For all three experiments, it was expected that stimulus familiarity differences would not appreciably influence IAT effects.

EXPERIMENT 1

Experiment 1 examined prejudice based on religious ethnicity (Jewish vs. Christian). The target concepts were Christian and Jewish names, matched on length and frequency in the population. These constraints were not expected to influence IAT effects. That is, both groups were expected to demonstrate implicit prejudice, given that ingroup preference is pervasive (Brewer, 1979) and anti-Semitism remains widespread in the United States (Smith, 1993). In contrast, scores on a self-report measure of anti-Semitism were not expected to differ between groups.

SUBJECTS

Thirty-six volunteers from introductory psychology courses at the University of Washington participated in exchange for course credit. Twenty-eight subjects were Christian (11 male, 17 female) and eight were Jewish (5 male, 3 female).[1] The representation of Jewish subjects in Experiment 1 is small, but typical of prejudice research using minority members (Fazio et al., 1995). Subjects participated individually in separate cubicles (up to three subjects participated simultaneously).

1. The original sample size was 39; three Christian subjects were excluded from analyses because their error rates indicated they were not attending to the task.

TABLE 1. Illustration of the IAT (Experiment 1)

	Left	Right
Condition 1		
Step 1	Jewish names	Christian names
Step 2	pleasant words	unpleasant words
Step 3	Jewish + pleasant	Christian + unpleasant
Step 4	unpleasant words	pleasant words
Step 5	Jewish + unpleasant	Christian + pleasant
Condition 2		
Step 1	Jewish names	Christian names
Step 2	unpleasant words	pleasant words
Step 3	Jewish + unpleasant	Christian + pleasant
Step 4	pleasant words	unpleasant words
Step 5	Jewish + pleasant	Christian + unpleasant

Note. Steps 3 and 5 are preceded by 20 practice trials each (not shown).

MATERIALS FOR IMPLICIT MEASURES

The Implicit Semitism IAT used 100 stimulus words: 25 Christian surnames (e.g., Higgins, Tyler), 25 Jewish surnames (e.g., Goldberg, Cohen), 25 pleasant-meaning words (e.g., rainbow, paradise), and 25 unpleasant-meaning words (e.g., vomit, murder). Christian and Jewish names were matched on frequency according to the 1990 American Census. The pleasant and unpleasant words were selected from norms reported by Bellezza, Greenwald, and Banaji (1986). (The appendix contains the stimuli for all three experiments.)

DESIGN OF THE IAT

The five steps of the IAT, described with Experiment 1's materials, are illustrated in Table 1. In Condition 1, these steps are as follows. (1) Subjects distinguish *target concepts* by pressing the right key for Christian names and the left key for Jewish names. (2) Subjects distinguish the *evaluative dimension* by pressing the right key for unpleasant words and the left key for pleasant words. (3) They respond to Jewish names and pleasant words with the left key and Christian names and unpleasant words with the right key (combined categorization task, abbreviated as Jewish + pleasant).[2] (4) They repeat Step 2 but with responses reversed (i.e., they press the right key for pleasant words, the left key for unpleasant words). (5) They respond to Jewish names and unpleasant words with the left key and Christian names and pleasant words with the right key (abbreviated as Christian + pleasant). The IAT effect is computed by

2. This abbreviation is arbitrary; the task could equally be described as Christian + unpleasant.

subtracting the mean response latency for performing the Jewish + pleasant task (Step 3) from the Christian + pleasant task (Step 5). Thus, positive difference scores reflect an automatic association between Jews and positive evaluation and Christians and negative evaluation (i.e., implicit Semitism). The order in which subjects perform Step 3 and Step 5 is counterbalanced across subjects (see Condition 2, Table 1). Nonorthogonally, key assignment for Step 2 is also counterbalanced. That is, subjects who perform the Christian + pleasant task first also press the right key for pleasant words and the left key for unpleasant words in Step 2 (see Condition 2, Table 1). In the present research, effects for the counterbalanced procedural variables were nonsignificant in each experiment (cf. Greenwald et al., 1998).

EXPLICIT MEASURES

Subjects completed self-report measures of attitudes and sociocultural background. There were three self-report attitude measures: a feeling thermometer, a semantic differential, and the Anti-Semitism Scale (Allport & Kramer, 1946). The thermometer measure asked subjects to indicate—separately for the social categories of Christians and Jews—how favorable each category was on a vertical scale labeled at 10 degree intervals from 0 (*very cold, or unfavorable*) to 99 (*very warm, or favorable*). The semantic differential consisted of five semantic differential items for Christians and Jews as separate concepts. These 7-point scales were anchored at either end by the polar-opposite adjectives, *beautiful-ugly, good-bad, pleasant-unpleasant, honest-dishonest,* and *nice-awful.* The semantic differential was scored by averaging the items for each object, scored on a scale ranging from 3 (positive) to -3 (negative). Difference scores were then computed for the thermometer and semantic differential measures such that higher scores reflect more positive evaluation of Jews compared to Christians. The Anti-Semitism Scale consists of 8 items scored on a 4-point scale ranging from 1 (agree) to 4 (disagree). Questions include "I can imagine myself marrying a Jewish person" and "One big problem with Jews is that they are never contented, but always try for the best jobs and the most money." The scale was scored and averaged so that high scores indicate pro-Semitic attitudes.

A measure of acquaintance differences was used to assess whether differential contact with outgroup members would influence implicit prejudice (Allport, 1954). The acquaintance measure asked subjects to provide the initials of up to 20 people that they knew, preferably close friends, but not family members. Subjects then indicated for each initial whether the person represented was *Jewish, Christian, None of the Above,* or *Don't Know.*

Scores for the acquaintance measure were converted to percentages, and a difference score computed such that positive scores reflect a higher percentage of Jewish acquaintances, relative to Christian.[3]

DESIGN AND PROCEDURE

The order in which subjects performed the Jewish + pleasant and Christian + pleasant combined tasks was counterbalanced. The design was a 2 (Subject Religion) × 2 (combined categorization task order) between-subjects factorial. After receiving computerized instructions, subjects began the IAT.[4] The experiment was administered on IBM-compatible desktop computers. Responses were assigned to the left and right forefingers (using the "A" and "5" keys on the numeric keypad, respectively). IAT stimuli appeared within a white window, vertically and horizontally centered against a light gray background. Subjects viewed this display from a distance of approximately 65 cm.

The stimuli were presented in blocks of 50 trials. Each trial block began with instructions describing the category discrimination(s) for the upcoming block and the assignment of response keys (left or right) to categories. Reminder labels, in the form of category names appropriately positioned to the left or right, remained on-screen during each block. Each combined category discrimination (e.g., Jewish + pleasant, Christian + unpleasant) consisted of a 50-trial practice block followed by two experimental blocks (total trials = 550 per subject). Subjects received summary feedback consisting of their mean response latency in ms and percent correct following each block. All blocks were subject-initiated (by pressing the space bar when ready). On each trial, the stimulus item was visible until the subject responded, and was replaced by a blank if the response was correct or by the word "error" if it was incorrect. Trials were conducted with a 250 ms interval between response to one stimulus and presentation of the next.

Following completion of the IAT, subjects responded to the explicit measures in privacy, under conditions designed to reduce self-presentation concerns. Subjects placed the completed measures

3. A similar measure of family members was also administered and scored, but proved to be almost redundant with subjects' religious classification, $r(34) = .96$, $p < .001$. In contrast, the acquaintance measure correlated less redundantly with subjects' religious classification, $r(34) = .59$, $p < .001$. We therefore retained only the acquaintance measure for further analyses.

4. The IAT programs used in the present research were written by Sean Draine and Shelly Farnham.

(marked only with their subject number) in a box containing several other measures to protect their anonymity.

RESULTS AND DISCUSSION

Data Reduction. The data for each trial block included response latencies (in ms) and error rates. To correct for anticipatory responses and momentary inattention, response latencies greater than 3000 ms and less than 300 ms were recoded as 3000 and 300 ms, respectively. The first two trials of each block were dropped because of their typically lengthened latencies. Latencies were log-transformed to employ a statistic that has satisfactory distribution of variance for analyses.[5] Analyses revealed relatively low error rates (an average of 8%), but were consistent with latency analyses in that higher error rates were obtained for conditions that produced longer latencies. As in previous research, error rates were less influenced than were latencies by task compatibility effects (Greenwald et al., 1998).

The IAT Effect. Figure 1 presents Experiment 1's results separately for Christian and Jewish subjects. Tasks combining Jewish names with pleasant concepts and Christian names with unpleasant concepts (abbreviated as Jewish + pleasant) are shown as black bars. Tasks combining Christian names with pleasant concepts and Jewish names with unpleasant concepts (abbreviated as Christian + pleasant) are shown as white bars. The expectation was that Christian subjects would show longer latencies when associating Jewish + pleasant versus Christian + pleasant, but that Jewish subjects would show the reverse. Figure 1 reveals this expected pattern (higher black bar vs. white for Christian subjects; higher white bar vs. black for Jewish subjects).

The IAT effect is computed as the difference in latencies between the two conditions. As shown in Figure 1, Jewish subjects averaged an IAT effect of +46 ms, whereas Christian subjects averaged a reverse IAT effect of -125 ms. Thus, as predicted, both groups showed superior performance when ingroup tokens were mapped onto pleasant words and outgroup tokens were mapped onto unpleasant words, compared to when these associations were reversed. An index of effect size (Cohen's *d*) was computed by dividing Jewish and Christian subjects' IAT effect scores by their pooled standard deviation. This computation yielded moderate and large effect sizes for Jewish (d = .48) and Christian (d = -1.31) subjects, respectively (Cohen, 1988).

5. Additional analyses were conducted on reciprocal conversions (i.e., 1000/latency in ms). All conclusions based on analyses of log-transformed latencies were equally evident on this measure.

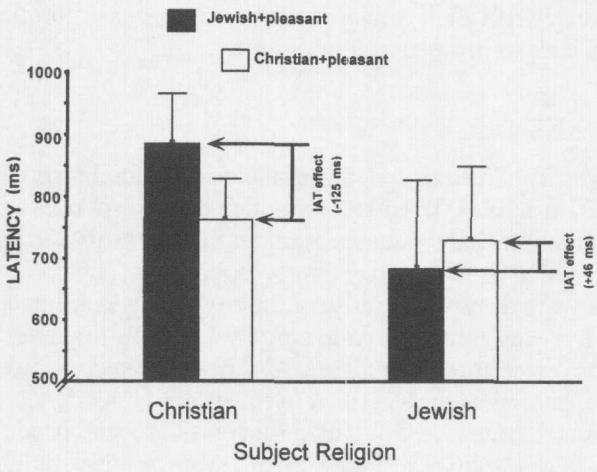

FIGURE 1. Mean latency results of Experiment 1 (N = 36) as a function of subject religion.

Only latencies for the combined categorization tasks are shown. Practice blocks and single categorization blocks are not included in the figure. Data are collapsed across the counterbalanced procedural variable, which did not significantly influence IAT effects. Error bars are 95% confidence intervals for the subjects contributing to each mean (28 Christians, 8 Jews).

The log latency IAT effect measure was submitted to a 2 (combined categorization task order) × 2 (subject religion) ANOVA. As expected, the implicit prejudice (IAT effect) was significant, $F(1, 32) = 14.53$, $p = .001$.[6] In addition, the main effect for subject religion was significant, $F(1, 32) = 21.96$, $p = 10^{-5}$.[7] The effects involving the counterbalanced procedural variables were nonsignificant, all Fs(1, 32) < 1.34, ps > .25.

IAT Compared with Explicit Measures. Table 2 presents summary data for Experiment 1's IAT and explicit measures, with all measures scored so that higher numbers were expected for Jewish subjects. The IAT, thermometer and semantic differential measures are difference scores computed so that positive numbers reflect more favorable attitudes toward Jews compared to Christians. As can be seen, the expected pattern emerged on the IAT, the feeling thermometer, and the semantic differential measure. In contrast, scores on the Anti-Semitism Scale were almost identical for both groups. However, the only explicit attitude measure to show significant group differences was the feeling thermometer (see Ta-

6.This test assesses whether IAT effect scores differed significantly from zero.

7. Following Greenwald, Gonzalez, Guthrie, and Harris (1996), p-values are reported as approximately exact values to provide information typically obscured by traditional p-value reports. Values less than .0001 are rounded to the nearest exponent of 10.

TABLE 2. Summary Statistics for Implicit and Explicit Measures (Experiment 1)

Measure	Jewish Subjects' Mean	Christian Subjects' Mean	Pooled *SD*	Group Differences (*t*)	*p*	Effect size (*d*)
Implicit Semitism IAT[a]	+46	-125	95.13	4.05	10^{-4}	1.80
Feeling Thermometer[b]	6.13	-11.21	18.43	2.34	.03	.94
Semantic Differential[b]	.20	-.36	.84	1.67	.10	.66
Anti-Semitism Scale[c]	3.58	3.54	.34	.27	.79	.12
Acquaintance[d]	-4%	-56%	30%	4.34	10^{-4}	1.73

Note. [a]Positive scores indicate more favorable implicit attitudes toward Jews, compared to Christians. [b]Positive scores indicate more favorable self-reported attitudes toward Jews, compared to Christians. [c]High scores indicate pro-Semitic attitudes. [d]Percentage of Jewish minus Christian acquaintances. Group differences are *t*-tests (df = 34) comparing Jewish and Christian subject means. Effect sizes (Cohen's *d*) were computed by dividing the Jewish minus Christian difference by the pooled *SD*. Conventional small, medium, and large effect sizes are .2, .5, and .8, respectively (Cohen, 1988).

ble 2). On the sociocultural measures, Jewish subjects reported a higher percentage of Jewish versus Christian family members, but their acquaintances were evenly spread between the two groups. Christian subjects reported a higher percentage of Christian versus Jewish acquaintances and family members. These group differences were significant (see Table 2).

Effect sizes for the difference between the two groups are also listed (computed by dividing the Christian minus Jewish difference by the pooled *SD*). Most notably, the IAT effect size shown in Table 2 ($d = 1.80$) was larger than those for the feeling thermometer ($d = .94$), semantic differential measure ($d = .66$), and the Pro-Semitism Scale ($d = .12$) combined. Thus, the implicit measure was more sensitive to group differences than these more conventional attitude measures.

The comparable effect sizes for the IAT and the thermometer measures suggested comparing their ability to classify subjects according to their religion. A logistic analysis in which subject religion was regressed hierarchically on the IAT and the thermometer measure revealed the IAT's superiority in this regard. The Wald statistic (a proxy for *F* in multiple regression) was 5.55 ($p = .02$) for the IAT, and .68 ($p = .41$) for the thermometer measure. Specifically, the IAT correctly classified 27 out of 28 Christian subjects, and 6 out of 8 Jewish subjects. The thermometer measure's correct classifications were 12 and 1, respectively. A scatterplot analysis revealed that the majority of subjects scored zero on the thermometer difference measure, whereas their implicit scores were in general below zero for Christians and above zero for Jews. The IAT's ability to correctly classify subjects who explicitly deny any preference

TABLE 3. Correlations Among Implicit and Explicit Measures (Experiment 1)

	IAT	Explicit Attitude Measures			Aquaintance Measure
Measure	1	2	3	4	5
1. Implicit Semitism IAT[a]					
2. Feeling Thermometer[a]	.34				
3. Semantic Differential[a]	.18	.83			
4. Anti-Semitism Scale[b]	.12	.51	.47		
5. Acquaintance[c]	.42	.36	.29	.13	
6. Subject Religion[d]	.63	.37	.28	.05	.60

Note. [a]Positive scores indicate more favorable attitudes toward Jews, compared to Christians. [b]High scores indicate pro-Semitic attitudes. [c]Percentage of Jewish minus Christian acquaintances. [d]Subject religion was coded 0 = Christian, 1 = Jewish. Correlations were computed using IAT log latency difference scores. Correlations with untransformed latencies were similar. $N = 36$ for all correlations; r values of .28, .33, .42, and .52 are associated with 2-tailed p-values of .10, .05, .01, and .001, respectively.

for one group over the other is a testament to its value in prejudice research (see also Greenwald et al., 1998).

Table 3 presents the correlations among Experiment 1's implicit and explicit measures. Measures were scored so that positive relationships were expected for all subjects. That is, on all measures high scores represent pro-Semitic attitudes. Subject religion was coded 0 = *Christian*, 1 = *Jewish*. As expected, the relations among Experiment 1's IAT and self-report attitude measures were generally weak. The only relationship to reach significance was that between the IAT and the thermometer measure, $r(34) = .34$, $p = .05$ (see also Greenwald et al., 1998, Experiment 2). The dissociation typically found between implicit and explicit measures of prejudice supports their conceptual distinction (Greenwald & Banaji, 1995). In Experiment 1, this dissociation was represented by the weak relations found between the implicit Semitism IAT and (a) the Anti-Semitism Scale (Allport & Kramer, 1946); and (b) the semantic differential.

As expected, the relationship between the acquaintance measure and the implicit Semitism IAT was reliable, $r(34) = .42$, $p = .01$. Thus, to the extent that subjects reported having more Jewish than Christian friends and acquaintances, they showed favorable implicit attitudes toward Jews. Although the acquaintance measure was also related to the thermometer measure, $r(34) = .37$, $p = 02$, it was unreliably related to the semantic differential, $r(34) = .29$, $p = 09$, and negligibly related to the Anti-Semitism Scale, $r(34) = .13$, $p = .44$. Finally, Table 3 shows the strong relationship between the IAT and subject religion. In sum, the IAT's relationships with subject religion and the acquaintance measure were ro-

bust, whereas the relationships shown between the IAT and direct measures of pro-Semitism were relatively modest.

EXPERIMENT 2

Experiment 1 showed that the IAT effect remained strong even when stimuli were matched on frequency in the language, and that the measure successfully classified subjects based on religion. Nonetheless, using census results to control for prior exposure may not have mirrored individuals' subjective sense of familiarity with Experiment 1's target concepts. Experiment 2 was conducted to further extend the generalizability of the IAT, but also to provide a stronger test of the effects of stimulus familiarity of the method. The focus of Experiment 2 was implicit ageism and age-related stereotypes. College-aged subjects completed two IATs, each using young (e.g., Kyle, Brittany) versus old (e.g., Ethel, Oscar) male and female names as the target concepts. The evaluative dimension for the *ageism* IAT consisted of pleasant and unpleasant words. The attribute dimension for the *age-stereotyping* IAT consisted of youthful (e.g., quick, sharp, bold) and elderly (e.g., slow, forgetful, cautious) stereotypic traits. Thus, Experiment 2 assessed negative stereotypes about older adults, and positive stereotypes about young people. Because Experiment 2 did not have an ingroup/outgroup design, the IAT was expected to reveal (on average) implicit ageism and age-related stereotypes, irrespective of subjects' scores on correspondent explicit measures (Perdue & Gurtman, 1990; Snyder & Miene, 1994). In addition, Experiment 2 directly assessed familiarity differences between target concepts to determine their influence on IAT effects.

METHOD

SUBJECTS

Fifty volunteers (23 male and 27 female) from introductory psychology courses at the University of Washington participated in exchange for course credit. From one to three subjects participated individually in separate cubicles.

MATERIALS FOR IMPLICIT MEASURES

The *ageism* IAT used 62 stimulus words: 16 names typical of young people (e.g., Tiffany, Ryan), 16 names typical of older people (e.g., Agnes, Clarence), 15 pleasant-meaning words (e.g., success, joy), and 15 unpleasant-meaning words (e.g., failure, poison). The *age-stereotype* IAT

used the same target concepts (young and old names), plus 10 attributes associated with youth (e.g., quick, sharp, bold) and 10 attributes associated with age (e.g., slow, forgetful, cautious). Young and old names were generated by the authors, with the aid of Internet web sites indicating popular baby names (for young names) and Civil War genealogy (for old names). Stereotypic attributes were adopted from past research (Purdue & Gurtman, 1990). (The appendix contains the stimuli used in Experiment 2.)

EXPLICIT MEASURES

Subjects completed a thermometer attitude measure, a semantic differential stereotyping measure, and a target concept familiarity measure. The thermometer measure assessed subjects' attitudes toward young and old people, as separate categories, on a vertical scale labeled at 10 degree intervals from 0 (*very cold*, or *unfavorable*) to 99 (*very warm*, or *favorable*). The stereotyping measure consisted of a set of seven semantic differential items for young and old people as separate categories. These 7-point scales were anchored at either end by the following polar-opposite adjectives: *slow-quick, forgetful-sharp, closed-open, frail-healthy, reserved-passionate, cautious-bold,* and *thrifty-generous.* The semantic differential was scored by averaging the items for each category, scored on a scale ranging from 3 (positive) to -3 (negative). Difference scores were then computed for the thermometer and semantic differential such that positive scores would reflect pro-youth attitudes and stereotypic judgments. The prior exposure measure asked subjects to indicate the extent to which they were familiar with each of the 32 names used as target concepts on a scale from 1 (*never heard* or *saw the name before this experiment*) to 7 (*extremely familiar*). A difference score was then computed such that positive scores would indicate more familiarity with young names relative to old names.

DESIGN

All subjects completed the ageism IAT and the age-stereotype IAT. The ageism IAT included youth-favorable (young + pleasant) and youth-unfavorable (old + pleasant) combined tasks. The stereotype IAT included stereotype compatible (young + youthful traits) and stereotype noncompatible (old + youthful traits) combined tasks. IAT order was counterbalanced. Order in which subjects completed tasks within each IAT was counterbalanced such that if young + pleasant preceded old+pleasant on the attitude IAT, young+youthful preceded old+youth-

ful on the stereotype IAT (and vice versa). The design was a 2 (IAT: ageism, age-stereotype) x 2 (IAT order) x 2 (combined categorization task order) x 2 (subject sex) mixed factorial, with repeated measures on the first factor.

PROCEDURE

Subjects completed the explicit measures before introduction to the IAT under the same conditions of anonymity used in Experiment 1. Administration of the IAT was similar to that of Experiment 1, except that subjects completed two IATs in succession. There were also minor programming changes. Single category discriminations consisted of 20 trials, versus 50. Each combined category discrimination consisted of a 20-trial practice block followed by a 40-trial experimental block (total trials = 360 per subject). Response accuracy feedback was displayed directly below the stimulus in the form of a green "O" for correct responses and a red "X" for incorrect responses. The red "X" remained on-screen until subjects corrected their response. Intertrial intervals were 150 ms following correct(ed) responses.[8]

RESULTS AND DISCUSSION

The IAT Effect. Figure 2 presents Experiment 2's results separately for the ageism and age-stereotype IATs. Data are collapsed across IAT order and task compatibility order. For the ageism IAT, subjects were expected to show longer latencies for tasks that combined old + pleasant compared to tasks that combined young + pleasant. Similarly, for the stereotype IAT, subjects were expected to show longer latencies for tasks combining old + youthful traits compared to tasks combining young + youthful traits. Figure 2 reveals this expected pattern (i.e., higher black bars than white bars for both IATs). The IAT effects shown in Figure 2 correspond to effect sizes of $d = 1.26$ and $d = 2.37$ for the ageism and age-stereotype IATs, respectively (see Table 4).

Ageism and age-stereotype log latency IAT-effect scores were submitted to a mixed factor analysis of variance (ANOVA). The IAT effect in this analysis was substantial, $F(1, 48) = 274.31$, $p = 10^{-21}$. This analysis also showed that the age-stereotype IAT yielded a stronger effect than

8. The program also had two additional features. First, it randomly presented stimuli from a list of possible words without replacement. Second, the program was constrained such that correct responses were not assigned to the same key more than three times in a row. However, these are not required features of the IAT.

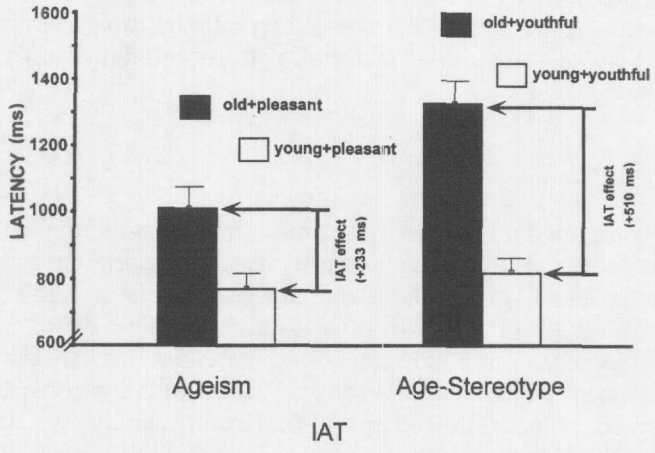

FIGURE 2. Mean latency results of Experiment 2 (N = 50) as a function of IAT.

Only latencies for combined categorization tasks are shown. Practice blocks and single categorization blocks are not included in the figure. Data are collapsed across counterbalanced procedural variables, which did not significantly influence IAT effects. Error bars are 95% confidence intervals for the subjects contributing to each mean (23 men, 27 women).

the ageism IAT, $F(1, 42) = 65.69$, $p = 10^{-10}$ (Ms = +510 vs. +233 ms). This finding may be due to the fact that the stereotype assessed was negative for the elderly, and positive for young people. Thus, this IAT combined specific beliefs and affect (i.e., was "doubly prejudicial" toward older

TABLE 4. Summary Statistics for Implicit and Explicit Measures (Experiment 2)

Measure	M	SD	Effect size (d)	Young-Old Category Differences t	p
Implicit					
Ageism IAT[a]	+233	184.42	1.26	8.92	10^{-12}
Age-Stereotype IAT[b]	+510	215.68	2.37	16.72	10^{-22}
Explicit					
Feeling Thermometer[a]	4.12	17.38	.24	1.66	.10
Semantic Differential[b]	2.33	.97	1.03	16.88	10^{-22}
Familiarity[c]	.38	1.30	.29	2.07	.04

Note. [a]High scores reflect more favorable attitudes toward young versus old people.
[b]High scores reflect more positive stereotyping of young versus old people.
[c]High scores reflect more familiarity with young versus old names.
Within subjects category difference scores were examined for their deviation from zero via t-tests (df = 49 for IAT variables, 48 otherwise). Effect sizes are Cohen's d. Effect sizes were computed by dividing mean difference scores by their standard deviations. Conventional small, medium, and large effect sizes for d are .2, .5, and .8, respectively.

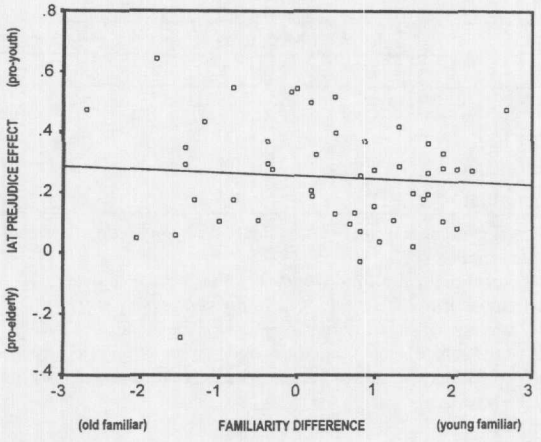

FIGURE 3. Logged IAT effect data of Experiment 2 ($N = 50$) as a function of familiarity differences between young and old names.

Positive IAT scores indicate implicit prejudice against old versus young people. Positive familiarity difference scores indicate more self-reported familiarity with young versus old names. The trend line is the regression slope for the effect of familiarity on IAT effects.

adults). The effects involving the counterbalanced procedural variables were nonsignificant, all Fs(1, 42) < 2.49, ps > .12.

A regression analysis was conducted to examine the effect of self-reported familiarity differences on logged IAT implicit prejudice scores. Figure 3 shows that the slope in this analysis was nonsignificant, indicating no effect of familiarity on these scores (beta = -.08, t = -.52, p = .61). In contrast, a significant intercept was shown, indicating a strong IAT effect in the absence of familiarity differences, $t(48) = 9.66$, $p = 10^{-13}$. An analysis examining the effect of familiarity differences on logged IAT implicit stereotype scores yielded similar results. Specifically, the slope was nonsignificant (beta = -.004, t = -.03, p = .98), whereas the intercept showed a significant IAT effect in the absence of familiarity differences, $t(48) = 18.03$, $p = 10^{-23}$.

IAT Compared with Explicit Attitude Measures. Table 4 presents summary data for the ageism and age-stereotype IATs along with the thermometer, semantic differential, and familiarity measures. All measures are difference scores wherein positive numbers reflect pro-youth attitudes and stereotypic judgments, and more familiarity with young versus old names.

The summary implicit data show the IAT effect scores for ageism and stereotyping (see also Figure 2) and their correspondent effect sizes (Cohen's d). Analyses of the direct measures showed a preference for young

TABLE 5. Zero-Order Correlations Among Implicit and Explicit Measures (Experiment 2)

	Implicit Measures		Explicit Measures	
Measure	1	2	3	4
1. Ageism IAT[a]	-.07	.41*	.28*	-.06
2. Age-Stereotype IAT[b]	.40*	-.00	.09	- .07
3. Feeling Thermometer[a]	.28*	.09	-.12	.22
4. Semantic Differential[b]	-.05	-.07	.22	-.12

Note. Correlations were computed using IAT log latency difference scores. Correlations with untransformed latencies were similar.

[a]High scores reflect more favorable attitudes toward young versus old people.

[b]High scores reflect more positive stereotyping of young versus old people.

Correlations below the diagonal are zero-order; correlations above the diagonal are partial, controlling for prior exposure; on the main diagonal, are the correlations between prior exposure and other measures. $N = 49$ for all correlations; r values of .23, .27, .35, and .44 are associated with 2-tailed p-values of .10, .05, .01, and .001, respectively.

versus old people on the feeling thermometer, resulting in a small effect size (d = .24).[9] The semantic differential measure showed greater endorsement of positive traits for young people compared to old people, resulting in a large effect size (d = 1.03). That is, subjects viewed young people as quicker, sharper, healthier, bolder, and more open, passionate, and generous than old people. However, the effect sizes on the self-report measures were smaller than those for the corresponding IAT measure. Specifically, the differences in effect sizes between the implicit and explicit measures were d = 1.02 for the attitude measures, and d = 1.34 for the stereotype measures. Finally, the familiarity measure showed that young names were rated as more familiar than were old names, resulting in a small effect size (d = .29).

Table 5 shows the correlations among Experiment 2's implicit and explicit measures. All measures were scored so that positive relationships might be expected. The table presents zero-order correlations below the diagonal, and partial correlations controlling for familiarity differences above the diagonal. Because the partial and zero-order correlations were comparable, only the latter will be discussed. On the main diagonal are correlations between the familiarity measure and other measures.

The main diagonal shows that differences between young and old target concept familiarity were unassociated with other dependent variables. The lower diagonal shows that implicit ageism and age-stereotyping were associated, $r(47)$ = .40, p = 004. In contrast, the self-report measures of attitude and stereotyping were unreliably re-

9. An outlier who scored -80.00 on the thermometer measure (z = -3.94) was eliminated from all self-report analyses.

lated, $r(47) = .22$, $p = .13$. This finding suggests that reliable links between stereotypes and prejudice may be found at the implicit level (see also Wittenbrink et al., 1997). Nonetheless, our measure of stereotyping included only negatively valenced attributes for the elderly and only positively valenced attributes for the young. Because this does not provide a measure of stereotyping unconfounded with prejudice, the relationship between the two measures may be due, in part, to their evaluative overlap.

Table 5 also shows that the relationships among the explicit and implicit measures were generally weak. Specifically, the ageism IAT and thermometer measures showed a small positive relationship, $r(47) = .28$, $p = .05$. That is, subjects who explicitly favored young people also showed evidence of implicit ageism. In contrast, the relationship between the semantic differential and age-stereotyping IAT was negligible, $r(47) = -.07$, $p = .63$. The dissociation among the implicit and explicit measures suggests that implicit ageism and stereotyping are independent of conscious beliefs (see also Perdue & Gurtman, 1990; Snyder & Miene, 1994).

In sum, Experiment 2 showed that the IAT is an effective measure of implicit age-related attitudes and stereotypes. The effect size for both measures was large (i.e., ds > 1.00). These IAT effects were not dependent on subjects' reporting significantly more familiarity for young versus old names, as shown by the regression analyses. These findings support the IAT's construct validity by suggesting that implicit prejudice and stereotyping effects are not significantly influenced by stimulus familiarity differences. This is important given the tendency for familiar stimuli to be preferred (Zajonc, 1968), and the likelihood that ingroup tokens will be more familiar than outgroup tokens in many intergroup relations investigations.

EXPERIMENT 3

Experiments 1 and 2 showed the IAT's usefulness when assessing implicit prejudice based on religion and age. They also suggested that IAT effects are not influenced by stimulus familiarity, with Experiment 2 providing the strongest test of this hypothesis. Finally, Experiment 2 showed that the IAT is easily configured to measure implicit stereotypes as well as attitudes (see also Rudman et al., 1998). Experiment 3 further extended the IAT's generality by assessing implicit prejudice based on nationality. The evaluative dimension consisted of pleasant and unpleasant words and target concepts were American versus Soviet leaders. To further investigate the effect of target concept familiarity differences on IAT-assessed prejudice, Experiment 3 directly manipu-

lated ingroup and outgroup familiarity. Specifically, target concept familiarity was counterbalanced in a design that crossed high and low familiarity with American and Soviet leaders. The objective was to provide, at minimum, two strongly differentiated conditions in which familiarity differences might be weak (unfamiliar American, familiar Soviet) versus robust (familiar American, unfamiliar Soviet).

A fully-crossed within subjects design would have required subjects to respond to four IATs (involving more than 1000 trials). To avoid subject fatigue, subjects responded to two IATs. Half the subjects responded to *matched-familiarity* IATs that (a) contrasted familiar American and Soviet leaders (e.g., Kennedy, Lincoln, Stalin, Lenin); and (b) contrasted unfamiliar American and Soviet leaders (e.g., Fillmore, Polk, Suslov, Mikoyan). The remaining half responded to *unmatched-familiarity* IATs that (c) contrasted familiar American and unfamiliar Soviet leaders (e.g., Kennedy, Lincoln, Suslov, Mikoyan); and (d) contrasted unfamiliar American and familiar Soviet leaders (e.g., Fillmore, Polk, Stalin, Lenin). Of particular interest was the comparison between conditions (c) and (d). These are conditions under which IAT effects might be strongest and weakest, respectively, if prior exposure effects contribute to implicit prejudice. It was expected that American subjects would reveal uniformly strong implicit prejudice against Soviet versus American leaders, irrespective of the familiarity combination manipulation.

METHOD

SUBJECTS

Eighty volunteers (30 male, 50 female) from introductory psychology courses at the University of Washington participated in exchange for course credit. From 1 to 4 students participated simultaneously in separate cubicles.

IMPLICIT AND EXPLICIT MATERIALS

The experiment's four IAT measures used 24 stimulus words: four familiar American leaders' names (Jefferson, Kennedy, Lincoln, Truman); four unfamiliar American leaders' names (Fillmore, Pierce, Polk, Tyler); four familiar Soviet leaders' names (Stalin, Lenin, Brezhnev, Khrushchev); four unfamiliar Soviet leaders' names (Mikoyan, Podgorny, Strogovich, Suslov); four pleasant-meaning words (gift, happy, health, miracle); and four unpleasant-meaning words (accident, divorce, poison, sickness). The American and Soviet leaders' names were generated by the authors on the basis of being judged to be clearly classifiable, yet

differentially familiar to the subject population. (The appendix contains the stimuli used in Experiment 3.)

Subjects also responded to a prior exposure measure on which they indicated the extent to which they were familiar with each of the 16 leaders on a scale from 1 (*never heard* or *saw the name before this experiment*) to 7 (*extremely familiar*). A difference score was then computed such that positive scores would indicate more familiarity with American versus Soviet names.

PROCEDURE AND DESIGN

Subjects first completed the prior exposure measure under the same conditions used in Experiments 1 and 2. Administration of the IAT was identical to that of Experiment 2. Subjects completed two IATs in which target concepts were either matched on familiarity (high American-high Soviet, low American-low Soviet) or unmatched on familiarity (high American-low Soviet, low-American, high-Soviet). In these IAT tasks, all subjects completed both compatible (American + pleasant, Soviet + unpleasant) and noncompatible (American + unpleasant, Soviet + pleasant) conditions. (This assumes that, for American subjects, it is appropriate to call the American + pleasant combination more evaluatively compatible than the Soviet + pleasant combination.) IAT order and task compatibility order were counterbalanced. The design was a 2 (familiarity combination: matched, unmatched) × 2 (familiarity order) × 2 (task compatibility order) between subjects factorial.

RESULTS AND DISCUSSION

Prior Exposure Differences. Within-subject contrasts were computed for all six possible pairwise comparisons between familiarity for American versus Soviet names. A Bonferroni procedure was used to provide a simultaneous .05 α-level for these tests. All pairwise differences were significant at the .05 level *except* for the comparison between the familiarity ratings of the familiar Soviet names and the unfamiliar American names; $t(74) = .44$, $p = .66$ (Ms= 4.37 and 4.28 respectively). Because these two item sets produced a nonsignificant familiarity difference, the IAT condition comparing these two sets will be of particular interest in the IAT analysis (below). If a large IAT effect is produced in this specific condition, then it is unlikely that the IAT effect is due solely to the familiarity disparity between items.

The IAT Effect. Figure 4 presents Experiment 3's results separately for the four IAT conditions representing the prior exposure manipulation.

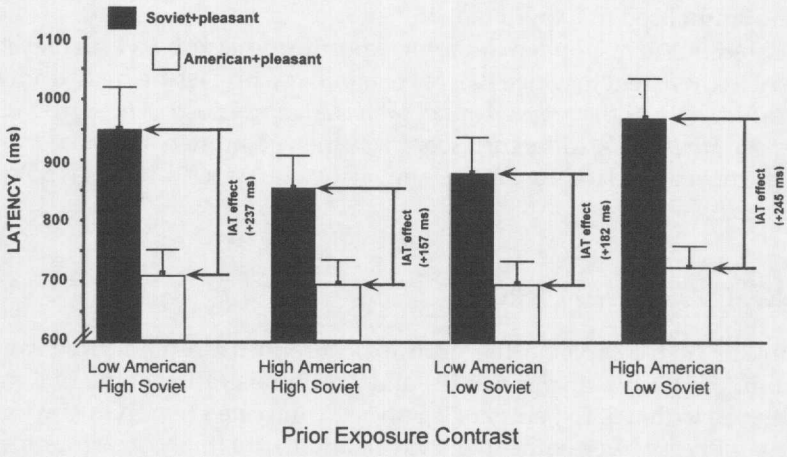

FIGURE 4. Mean latency results of Experiment 3 (N = 80) as a function of IAT.

Only latencies for combined categorization tasks are shown. Practice blocks and single categorization blocks are not included in the figure. Data are collapsed across counterbalanced procedural variables, which did not significantly influence IAT effects. Error bars are 95% confidence intervals for the subjects contributing to each mean (30 men, 50 women).

For each IAT, subjects were expected to show longer latencies for tasks that combined Soviet + pleasant, compared to tasks that combined American + pleasant. Figure 4 reveals this expected pattern (i.e., higher black bars than white bars for all conditions). The arrangement of conditions in Figure 4 from left to right represents an increase in self-reported familiarity disparity from smallest (Low American, High Soviet) to greatest (High American, Low Soviet). As can be seen, the IAT effect in the former condition (+237 ms) is nearly identical to that in the latter condition (+242 ms). Thus, the IAT effect appears to be independent of prior exposure differences for target concepts. Consistent with this view, an omnibus ANOVA comparing IAT effects across all four conditions was nonsignificant, $F(3, 71) = 2.25$, $p = .09$.

Because each subject participated in only two of the four familiarity combinations, analyses were conducted separately for each subject group (matched vs. unmatched familiarity) in order to analyze familiarity combination as a within-subjects factor. For each subject group, log latency IAT effect scores were submitted to 2 (familiarity combination) × 2 (IAT order) × 2 (task compatibility order) ANOVAs with repeated measures on the first factor. The Bonferroni method was again used to maintain a .05 α-level for the 16 possible effects. Only the tests for the difference of

TABLE 6. Regression Results for Target Concept Familiarity Combinations (Experiment 3)

Familiarity Combination	B	SE	Beta	t	p	R^2
Low American–High Soviet						
IAT effect (intercept)	.27	.03		9.02	10^{-11}	
Familiarity effect (slope)	-.007	.02	-.09	-.51	.61	.007
Low American–Low Soviet						
IAT effect (intercept)	.17	.05		3.53	.001	
Familiarity effect (slope)	.008	.02	-.07	.45	.65	.008
High American–High Soviet						
IAT effect (intercept)	.22	.04		5.60	10^{-6}	
Familiarity effect (slope)	-.004	.01	.08	-.39	.70	.003
High American–Low Soviet						
IAT effect (intercept)	.25	.05		5.26	10^{-6}	
Familiarity effect (slope)	.003	.01	.05	.32	.75	.004

Note. The dependent variable in each equation was the logged IAT difference score. High scores on this measure reflect implicit prejudice against Soviet versus American leaders. The predictor variable in each equation was the difference between self-reported familiarity for American versus Soviet names in each condition. High scores on this measure reflect more familiarity for American versus Soviet names. Matched (high-high, low-low) and unmatched (low-high, high-low) equations contain the same subjects ($N = 40$ for each).

the IAT effects from zero were significant, $F(1, 36) = 151.26$, and $F(1, 36) = 157.45$, $ps = 10^{-14}$ for the matched and unmatched familiarity groups, respectively. Effects of familiarity combination, familiarity order, and task compatibility order were all nonsignificant at the .05 level.

As in Experiment 2, regression analyses were conducted to test for the effect of self-reported familiarity differences on logged IAT prejudice scores. Table 6 shows the results of these analyses for each familiarity combination. The arrangement of conditions in Table 6 from top to bottom represents an increase in self-reported familiarity disparity from smallest (Low American, High Soviet) to greatest (High American, Low Soviet). In each equation, the intercept for the IAT effect was significant, whereas the regression coefficient for familiarity was nonsignificant. In each condition, the amount of IAT effect variance accounted for by familiarity differences was less than .01. These analyses provide further evidence that IAT effects are independent of familiarity differences for target concepts.

In Experiment 3, implicit prejudice based on nationality was expected and shown, irrespective of the prior exposure manipulation. Subjects performed faster on American + pleasant tasks than on Soviet + pleasant tasks in all four familiarity combinations. Using the pooled standard deviation (for compatible and noncompatible conditions) as the effect size unit and collapsing across all design factors other than familiarity combination, effect sizes for the IAT effects were $d = .98$ for only Soviet familiar (American unfamiliar), $d = .77$ for both familiar, $d = .85$ for both

unfamiliar, and $d = 1.08$ for only American familiar (Soviet unfamiliar). Most notably, effect sizes in the first and last conditions were essentially equivalent, even though these represented conditions in which the familiarity discrepancies were nonsignificant and highly significant, respectively. This finding, in tandem with the regression analyses, further substantiates interpreting the IAT effect as a measure of the implicit associations underlying attitudes rather than implicit preference for familiar stimuli.

GENERAL DISCUSSION

Across three experiments, the IAT effectively assessed implicit prejudice across a wide range of social groups. The specific prejudices examined were based on religious ethnicity (Jewish vs. Christian), age (young vs. old), and nationality (American vs. Soviet). In each case, the IAT revealed strong evidence for implicit prejudice. The average implicit prejudice effect size was $d = 1.32$, conventionally regarded as a large effect (Cohen, 1988). In contrast, the average effect size for self-report measures of prejudice was moderate, $d = .49$. This difference in effect sizes may be partly due to the greater number of items used in the IAT, compared to the thermometer measure, for example. In Experiment 2, the IAT also revealed strong evidence for implicit age-related stereotypes. The implicit stereotype effect size was more than twice that yielded by the explicit measure (ds = 2.37 vs. 1.03). The potency of IAT effect sizes and the ease with which the method is configured to assess multiple attitudes and stereotypes attest to its value as a flexible, user-friendly tool for investigating implicit social cognition (Greenwald et al., 1998; Rudman et al., 1998).

THE (NON)EFFECTS OF PRIOR EXPOSURE

The present research provides the first known tests of the influence of stimulus familiarity differences on implicit prejudice assessment. These tests are important given the likelihood of familiarity differences between ingroup and outgroup tokens and the influence of prior exposure on attitudes (Zajonc, 1968). Experiment 1 showed that Christian and Jewish subjects favored their own group, even though target concepts were matched on population frequency. Experiment 2 revealed that subjects possessed implicit pro-youth attitudes and stereotypes irrespective of reported familiarity differences for young versus old names. Experiment 3 deliberately manipulated the familiarity of target concepts, affording a test of the IAT effect under conditions representing a range of familiarity differences from nonsignificant to highly significant. Results showed implicit prejudice against Soviet versus American leaders in all

conditions, irrespective of stimulus familiarity. In combination, the findings provide persuasive evidence that the IAT measures implicit prejudice independent of preference for familiar stimuli. Of course, this does not preclude the possibility that implicit attitudes will be affected by actual contact with outgroup members (Allport, 1954). In this respect, Experiment 1's results were suggestive: Compared to Christians, Jewish subjects showed both higher percentages of outgroup acquaintances and lower implicit prejudice.

IMPLICIT VERSUS EXPLICIT DIFFERENCES

Consistent with prior research (e.g., Greenwald et al., 1998; Rudman et al., 1998), the IAT and self-report measures of Experiments 1 and 2 were weakly or unreliably related to one another, indicating that the two types of methods assess independent constructs (Greenwald & Banaji, 1995). In Experiment 1, the extent to which subjects showed implicit ingroup + pleasant associations was generally unrelated to explicit measures of prejudice (average $r = .21, p = .22$). In Experiment 2, implicit and explicit ageism measures were modestly related ($r = .28, p = .05$) and the age-stereotyping counterparts were unreliably related ($r = -.07$, ns). Given these weak links, it is noteworthy that Experiment 1's IAT showed stronger evidence of known groups validity than did conventional measures of attitudes (see also Greenwald et al., 1998, Experiment 2).

IMPLICATIONS OF THE RESEARCH

Whether subjects knowingly misrepresent their attitudes on self-report measures or are genuinely unaware of the discrepancy between their implicit and explicit attitudes, the mounting evidence for pervasive implicit stereotyping and prejudice suggests a social reality in which automatic ingroup favoritism may be an inescapable fact (e.g., Fazio et al., 1995; Dovidio et al., 1997; Greenwald et al., 1998). The present research adds to this evidence, which has implications for policies designed to ensure equal opportunity (e.g., affirmative action programs) and for the backlash against them in recent times (Greenwald & Banaji, 1995). The preference for self, and by extension, entities connected to the self, may be irresistible (Greenwald, Banaji, Rudman, Farnham, Nosek, & Rosier , in press; Perdue et al., 1990). This inherent bias may best be offset by reclassifying outgroup members as ingroup members (e.g., Dovidio, Gaertner, Validzic, & Matoka, 1997; Gaertner, Mann, Murrel, & Dovidio, 1989). However, in the absence of reclassification, external policies provide important insurance for minorities and other disenfranchised groups. Moreover, irrespective of the specific approach, efforts to combat prejudice and stereotyping can only be improved (and their effec-

tiveness properly examined) by the use of assessment tools that are capable of measuring implicit attitudes and beliefs in a wide variety of domains. The present research, in tandem with prior evidence (e.g., Greenwald et al., 1998; Rudman et al., 1998), shows that the IAT effectively serves this need.

CONCLUSION

Three experiments showed that the IAT is a powerful and flexible measure of implicit prejudices based on religion, age, and nationality. Results supported the method's construct validity by extending its known groups validity (Experiment 1) and by showing that the IAT assesses implicit prejudice independent of self-reported stimulus familiarity differences (Experiments 2-3). They also showed that the IAT can be easily configured to measure implicit stereotypes (Experiment 2). The IAT's flexibility and generality affords researchers the opportunity to expand prejudice and stereotype assessment to new horizons, and to overcome the obvious limitations of self-report methods.

APPENDIX. Stimuli for Three Experiments

Experiment 1

Unpleasant Words	Pleasant Words	Christian Names	Jewish Names
evil	honor	Bankhead	Birnbaum
cancer	lucky	Abernathy	Blumenthal
sickness	diamond	Higgins	Cohen
disaster	loyal	Gerhardt	Edelman
poverty	freedom	Hazelwood	Eisenberg
vomit	rainbow	Holloway	Schwartz
bomb	love	Delancey	Fleischman
rotten	honest	Buckley	Jacobson
abuse	peace	Bingham	Goldberg
murder	heaven	Elkins	Katz
assault	pleasure	Tyler	Klein
grief	family	Everson	Lieberman
divorce	diploma	Holcomb	Levine
poison	gift	Lyles	Rosen
kill	cheer	Watterson	Rothstein
death	health	Millet	Saltzman
hatred	friend	Andersen	Friedman
ugly	caress	Ledford	Siegel
accident	gentle	Carruthers	Silverstein
jail	happy	Crowell	Shapiro
stink	miracle	Duffy	Levy
tragedy	sunrise	Winstead	Weinstein
crash	paradise	Copeland	Zimmerman
filth	vacation	Harkness	Gottlieb
pollute	laughter	Barker	Zucker

APPENDIX. Stimuli for Three Experiments (continued)

Experiment 2

Pleasant Words	Unpleasant Words	Stereotypic Young Attributes	Stereotypic Old Attributes	Young Names	Old Names
caress	abuse	healthy	frail	TIffany	Ethel
cuddle	agony	quick	slow	Christine	Bernice
glory	assault	sharp	forgetful	Julia	Beverly
gold	brutal	flexible	rigid	Brianna	Lucille
health	corpse	curious	irritable	Jason	Cecil
joy	death	bold	cautious	Justin	Myron
kindness	failure	open	closed	Alex	Vernon
lucky	filth	passion	reserved	Kyle	Wilbert
peace	killer	nimble	stodgy	Brittany	Gertrude
snuggle	poison	generous	thrifty	Kelsey	Agnes
success	slime			Danielle	Winnifred
sunrise	slum			Gillian	Adelaide
talent	stink			Ryan	Clarence
triumph	torture			Cameron	Irwin
warmth	vomit			Brandon	Oscar
				Corey	Alfred

Experiment 3

Pleasant Words	Unpleasant Words	Familiar American Leaders	Unfamiliar American Leaders	Familiar Soviet Leaders	Unfamiliar Soviet Leaders
health	divorce	Jefferson	Tyler	Khrushchev	Podgorny
happy	sickness	Lincoln	Polk	Brezhnev	Suslov
miracle	accident	Truman	Fillmore	Lenin	Mikoyan
gift	poison	Kennedy	Pierce	Stalin	Strogovich

REFERENCES

Allport, G. W. (1954). *The nature of prejudice.* New York: Addison-Wesley.

Allport, G. W., & Kramer (1946). Some roots of prejudice. *Journal of Psychology, 22,* 9–39.

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, & Computers, 18,* 299–303.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86,* 307–324.

Cohen, J. (1988). *Statistical power for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In Tanur, J. M. (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.

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

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.

Duckitt, J. H. (1992). Psychology and prejudice: A historical analysis and integrative framework. *American Psychologist, 47,* 1182–1193.

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

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57,* 239–249.

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., & Rosier, M. (in press). Prologue to a unified theory of attitudes, stereotypes, and self-concept. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition and behavior.* New York: Cambridge University Press.

Greenwald, A. G., & Breckler, S. J. (1985). To whom is the self presented? In B. R. Schlenker (Ed.), *The self and social life* (pp. 126–145). New York: McGraw-Hill.

Greenwald, A G., Gonzales, R., Guthrie, D. G., & Harris, R. J. (1996). Effect sizes and *p*-values: What should be reported and what should be replicated? *Psychophysiology, 33,* 175–183.

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.

Jones, J. M. (1972). *Prejudice and racism.* New York: McGraw-Hill.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76,* 349–364.

Judd, C. M., Park, B., Ryan, C. S., Brauer, M., & Kraus, S. (1995). Stereotypes and ethnocentrism: Interethnic perceptions of African American and White American college samples. *Journal of Personality and Social Psychology, 69,* 460–481.

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–126). Orlando, FL: Academic Press.

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

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26,* 199–216.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (1998). *Sex differences in gender stereotypes revealed by the Implicit Association Test.* Manuscript submitted for publication.

Rudman, L. A., & Kilianski, S. E. (in press). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin.*

Selznick, G. J., & Steinberg, S. (1969). *The tenacity of prejudice: Anti-Semitism in contemporary America.* New York: Harper and Row.

Smith, T. W. (1993). Actual trends or measurement artifacts? A review of three studies of anti-Semitism. *Public Opinion Quarterly, 57,* 380–393.

Sniderman, P. M., & Carmines, E. G. (1997). *Reaching beyond race.* Cambridge, MA: Harvard University Press.

Snyder, M., & Miene, P. K. (1994). Stereotyping of the elderly: A functional approach. *British Journal of Social Psychology, 33,* 63–82.

Spence, J. T., & Hahn, E. D. (1997). The Attitudes Toward Women Scale and attitude change in college students. *Psychology of Women Quarterly, 21*(1), 17–34.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(Supp. 2, Pt. 2).