

Single
EM = .914
(diff. confounding)

Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes

Allen R. McConnell

Miami University

and

Jill M. Leibold

Michigan State University

Received June 16, 2000; revised November 17, 2000; accepted November 17, 2000; published online June 8, 2001

10
Heretofore, no research has shown that meaningful variability on the Implicit Association Test (IAT) relates to intergroup discrimination or to explicit measures of prejudice. In the current study, White undergraduates interacted separately with White and Black experimenters, and their behavior during these social interactions was assessed by trained judges and by the experimenters themselves. The participants also completed explicit measures of racial prejudice and a race IAT. As predicted, those who revealed stronger negative attitudes toward Blacks (vs Whites) on the IAT had more negative social interactions with a Black (vs a White) experimenter and reported relatively more negative Black prejudices on explicit measures. The implications of these results for the IAT and its relations to intergroup discrimination and to explicit measures of attitudes are discussed. © 2001 Academic Press

32

Since LaPiere's (1934) classic demonstration of attitude-behavior inconsistency toward a Chinese couple traveling across the United States, social psychologists have invested a great deal of energy into developing techniques to assess group attitudes in ways that circumvent problems resulting from limited introspective access, experimenter effects, and social desirability concerns. Recently, researchers have employed various social cognition approaches to assess prejudice that minimize the problems involved with explicit reports of attitudes (e.g., Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Wittenbrink, Judd, & Park, 1997). The

current work focuses on the most recent of these techniques, the Implicit Association Test (IAT), to examine the extent to which it relates to intergroup behavior and to explicit measures of racial attitudes.

The IAT has become a widely used instrument to measure attitudes in general, and prejudices toward groups in particular. It assesses attitudes by having people quickly categorize stimulus words using two response keys. In racial IAT studies, the stimulus words are names that are racially stereotyped (e.g., Jamal and Sue Ellen) or adjectives that have evaluative connotations (e.g., wonderful and disgusting). In critical trial blocks, participants categorize these words using two keys, each of which has two response options mapped to it. Typically, White participants categorize the words more quickly when "Black or undesirable" is mapped onto one key response and "White or desirable" is mapped onto the other key response than when the opposite set of key mappings (i.e., "Black or desirable" and "White or undesirable") are used (Greenwald et al., 1998). The difference in the average response latency between these two sets of key mappings is known as the IAT effect. Presumably, larger IAT effects reflect stronger associations in memory between the concept pairings (i.e., those re-

This research was supported by NIMH Grants MH58449 and MH60645. The authors gratefully acknowledge the outstanding contributions of Ayana Burroughs, Kristy Dean, and Amy Johnson, who served admirably as experimenters in this study. The diligent efforts of Ben Diaz and Zachary Eichenhorn in coding videotapes are also greatly appreciated. The helpful comments of Frank Bernieri, Chris End, and two anonymous reviewers on a previous draft of this article are greatly appreciated.

Address correspondence and reprint requests to Allen R. McConnell, Department of Psychology, Miami University, Oxford, OH 45056-1601. E-mail: mcconnar@muohio.edu.

sponses that shared the same response key) that facilitated judgment.

Social psychologists who study group prejudice have been drawn to the IAT because of its large effect size (Greenwald et al., 1998; Nosek, Banaji, & Greenwald, 2000) and because even people who know that the IAT assesses group prejudice still reliably produce the IAT effect, indicating its robustness and apparent imperviousness. As a result, the IAT appears to circumvent many of the problems of traditional, explicit measures of prejudice. Researcher enthusiasm and large effect sizes notwithstanding, the IAT effect has not been demonstrated to be related to behavior toward group members. Although strong between-group differences have revealed favoritism for religious, ethnic, age-related, and racial ingroups (Greenwald et al., 1998; Rudman, Greenwald, Mellott, & Schwartz, 1999), meaningful variability in the strength of the IAT effect has not been shown to be related to one's discriminatory behavior.

Other implicit measures of prejudice, for example, have been shown to relate to intergroup behavior (Dovidio et al., 1997; Fazio et al., 1995). Indeed, the current research married the methodologies of this previous research to examine whether the IAT predicts intergroup discrimination. Thus, the primary goal of the current study was to explore whether this relation exists, which would help substantiate the predictive utility of the IAT.

In addition, the current work also examined whether the IAT relates to explicit measures of prejudice. It has been argued that implicit and explicit measures of attitudes tap into different knowledge and thus should be unrelated (Greenwald & Banaji, 1995; Greenwald et al., 1998), whereas others have found relations between the two (Dovidio et al., 1997, Experiment 2; Wittenbrink et al., 1997; cf., Dovidio et al., 1997, Experiment 3). To the extent that explicit measures of prejudice are reactive and subject to normative pressures, a lack of correspondence between implicit and explicit measures of prejudice is not surprising (Dunton & Fazio, 1997; Fazio et al., 1995). Although other factors, such as desire to avoid discriminatory responses because they are inconsistent with one's values (e.g., Plant & Devine, 1998), can influence behavior toward group members, minimizing self-presentation concerns should, at least, increase the likelihood of observing attitude-behavior consistency (Fazio, 1990). With respect to the IAT, Greenwald et al. (1998) did not find a correlation between the IAT and explicit measures of prejudice (i.e., feeling thermometer and semantic differential scales). The current work examined whether a relation between the IAT and explicit measures would be revealed under conditions designed to minimize self-presentation concerns.

In sum, the current study explored the relations among the IAT, intergroup behavior, and explicit reports of prejudice. Participants met with a White experimenter, and later

with a Black experimenter, in structured social interactions. These interactions were videotaped and later assessed by trained judges. Also, the Black and White experimenters independently assessed their interaction during the course of the experiment. Thus, a within-subjects design allowed us to examine how each participant behaved toward a Black experimenter relative to a White experimenter. Before interacting with the Black experimenter, participants privately completed a series of questionnaires to assess their attitudes toward Blacks and Whites in a minimally reactive situation. Afterward, they completed a race IAT before having an unanticipated social interaction with a Black experimenter. It was predicted that those who revealed relatively more negative attitudes toward Blacks on the IAT would behave in a relatively less friendly fashion toward the Black experimenter. This finding would substantiate the predictive validity of the IAT and suggest that it assesses individuals' idiosyncratic attitudes.

Two other empirical questions were also examined. First, would the IAT relate to explicit reports of prejudice? Previous research on implicit measures has yielded mixed results. Second, would explicit reports of prejudice relate to behavior toward the Black experimenter? It was our belief that the likelihood of observing significant relations between explicit measures of prejudice and other outcomes (i.e., IAT, behavior) would be improved under conditions in which participants felt minimal presentational concerns.

METHOD

Participants

At Michigan State University, 42 White undergraduates enrolled in introductory psychology courses participated in exchange for extra credit.

Measures

Explicit measures of prejudice. Participants completed semantic differential scales for Blacks, semantic differential scales for Whites, a feeling thermometer for Blacks, and a feeling thermometer for Whites (in that order). Each measure was completed on a separate page in a questionnaire booklet. Seven-point scales were used for the semantic differential word pairings: beautiful-ugly, good-bad, pleasant-unpleasant, honest-dishonest, and nice-awful. Participants also reported their attitudes toward Blacks and Whites using a feeling thermometer, which ranged from 0° (*extremely unfavorable*) to 100° (*extremely favorable*).

IAT task. Participants completed a word-based IAT task, which presented 96 stimulus words: 24 Black-associated names (e.g., Jamal and Yolanda), 24 White-associated names (e.g., Fred and Mary Ann), 24 desirable words (e.g., wonderful and awesome), and 24 undesirable words (e.g., offensive and disgusting). Names were always presented in

EMs
13
27
30
17
18
26
29
31
33

TABLE 1
Trial Blocks Used in the IAT Task

Block(s)	Type of judgment	Left key	Right key
1	Name discrimination	Black	White
2	Adjective discrimination	Undesirable	Desirable
3 and 4	Prejudice consistent combination	Black or Undesirable	White or Desirable
5	Reversed name discrimination	White	Black
6 and 7	Prejudice inconsistent combination	White or Undesirable	Black or Desirable

Note. Left key refers to categories associated with the "D" response, and right key refers to categories associated with the "K" key response.

uppercase letters, and adjectives were always presented in lowercase letters.

The IAT task was based on Greenwald et al. (1998), using a computer program written by the first author. As Table 1 reports, participants encountered five types of trial blocks across seven different blocks, with each block being composed of 48 trials. For half of the participants, Blocks 3 and 4 presented the prejudice-inconsistent combination and Blocks 6 and 7 presented the prejudice-consistent combination (the left key and right key response options for Blocks 1 and 5 were also reversed). This block order manipulation did not produce any effects and thus receives no further discussion. In Blocks 1, 2, and 5, each of the 48 relevant stimulus words was presented once based on a randomly determined order. In each of the combination blocks, the word types were alternated across trials (i.e., name, adjective, name, adjective, and so forth) with individual stimulus words selected at random from their respective lists until each of the 48 relevant items had been presented once across the two blocks.

Participants were told that they would be making a series of category judgments. On each trial, a stimulus word was displayed in the center of a computer window (24-point black serif text on a gray background), and participants used the "D" or "K" key on the keyboard for their responses. Category label reminders were displayed in blue text on the left and right sides of the window. Participants were told, "Make your judgments as rapidly as possible, but don't respond so fast that you make many errors. Occasional errors are okay. If you do make a mistake, a red X will appear on the screen below the target word. Please press the correct category key to continue. You cannot continue until you make the correct response." Participants were told to keep their index fingers on the "D" and "K" keys throughout the experiment to minimize delays in responding. A 250-ms gray-screen intertrial interval was used. In between blocks, participants were given a self-paced break and instructions for the next block.

Procedure

Participants arrived at the laboratory for an experiment on "word perception" and were greeted by a White female experimenter.¹ They were run individually. Unbeknown to the participant, a hidden video camera was positioned to record the participants' and experimenters' full bodies and their entire range of movements during scripted social interactions. A hidden unidirectional microphone recorded their discussions. They were directed to a rolling desk chair initially positioned 120 cm away from the experimenter's chair, allowing participants to establish a preferred distance from the experimenter. The experimenter explained that because the experiment was brief, the participant would complete four unrelated tasks. For the first task, they were told that the Department of Psychology had asked experimenters to interview students about their experiences in psychology. The experimenter asked the participant four innocuous questions (e.g., "What would you change to improve psychology classes?"), pausing for the participant's response between each question and recording the responses on a report form. The experimenter also told a scripted joke following the second question. This interaction took about 3 min.

Next, participants completed a booklet of questionnaires that purportedly were being used to develop future experiments. They were told that it was important for them to answer honestly in order for the future research projects to be successful. The privacy of their responses was stressed by explaining that they would complete the booklet in a private room, place the completed booklet in a sealed envelope, and drop it into a covered box without any experimenter interaction. The booklet contained several questionnaires, only some of which were relevant to the current study. After completing several pages of the booklet, participants completed the semantic differential scales and the feeling thermometer measures. It took participants about 15–20 min to complete the booklet.

While the participant was completing the booklet, the White experimenter assessed her interaction with the participant (details forthcoming). After completing the booklet, participants inserted the sealed survey into a covered box in the laboratory's waiting area. They then found the experimenter, who took them to a private computer workstation to

¹ The sequence of events that participants experienced in the experiment was fixed to minimize suspicion about the overall goals of the study (e.g., initially encountering a Black experimenter may have raised immediate concerns that the study was about racism). Although it is possible that exposure to one's own responses on the IAT or the explicit prejudice measures might affect subsequent behavior toward the Black experimenter, we reasoned that because the interaction with the Black experimenter was unexpected, participants would find it difficult to control their subtle behavioral cues toward her in an extemporaneous social interaction. However, we acknowledge that a fixed-order design may introduce the possibility of unforeseen confounds in the current study.

begin the "word perception" experiment (i.e., the IAT). The experimenter then looked at the clock and mentioned that her shift was almost over and that a new experimenter would assist in completing the fourth task following the word perception experiment. At that point, the White experimenter started the IAT program and excused herself. Participants required about 10 min to complete the IAT.

While participants completed the IAT, a Black female experimenter replaced the White experimenter and greeted participants after they returned from their room after completing the IAT. Once again, the participant was directed to a chair positioned 120 cm from the experimenter's chair, allowing the participant to establish a preferred seating distance. The Black experimenter asked the participant seven questions about the experiment (e.g., "What did you think about the difficulty level of the computer task?" and "Were the instructions clear?"), pausing for the participant's response between each question and recording the responses on an interview form. She also told a scripted joke after the fourth question. Afterward, the experimenter explained that both social interactions had been videotaped, and she asked for the participant's permission to use the videotape for data analyses. One participant refused, and her videotape was erased in her presence, leaving 41 participants for data analyses. Finally, participants were debriefed and thanked for their participation.

Coding of Social Interactions

Trained judges' ratings of participants' behavior. Based on the existing literature documenting behavior cues that convey emotions and attitudes (Crosby, Bromley, & Saxe, 1980; DePaulo, 1992; DePaulo & Friedman, 1998; Duncan, 1969; Eckman & Friesen, 1967; Hendricks & Bootzin, 1976; Kleinke, 1986; Word, Zanna, & Cooper, 1974), 16 behaviors were coded by two trained judges who were unaware of participants' attitudes. Using a scale from 1 (*none*) to 9 (*very much*), judges rated the participant's friendliness during the interaction, the abruptness or curtness of the participant's responses to questions, the participant's general comfort level, how much the participant laughed at the experimenter's joke, and the amount of participant's eye contact with the experimenter. On 5-point scales, they assessed the participant's forward body lean toward the experimenter (vs leaning away), the extent to which the participant's body faced the experimenter (vs facing away), the openness of the participant's arms (vs crossed arms), and the expressiveness of the participant's arms (vs not moving at all). Judges also calculated the distance between the experimenter and the participant's chair at the end of the interaction to gauge social distance. Judges also recorded the participant's speaking time, number of smiles, number of speech errors, number of speech hesitations (e.g., "um"), number of fidgeting body movements (e.g., swinging feet and shifting positions), and num-

ber of extemporaneous social comments made by the participant.² The judges rated each participant's interaction with the White experimenter and with the Black experimenter separately. The videotape showed both the participant and the experimenter, and the judges were instructed to only attend to the audio for ratings associated with the interaction dialogue (e.g., curtness of responses).

Experimenters' ratings. Each experimenter completed a 5-item inventory after their interaction with the participant. Using a scale ranging from 1 (*not at all*) to 9 (*extremely*), experimenters recorded their assessment of the participant's degree of eye contact, the abruptness or curtness of the participant's responses, the participant's friendliness, the participant's perceived comfort level during the interaction, and the experimenter's own comfort level during the interaction.

RESULTS

Data Reduction

Overview. The data analytic strategy was to transform all measures, implicit and explicit, into difference scores that reflected the relative degree of prejudice against Blacks (i.e., relatively more positive attitudes toward Whites than Blacks and relatively more positive behaviors toward Whites than Blacks). Thus for *all measures*, larger positive scores reflected greater negativity toward Blacks than Whites.

IAT. To reduce the positive skew inherent in response latency data (Greenwald et al., 1998; Ratcliff, 1993), a log transformation was applied to each response latency. IAT effect scores were computed by comparing mean response latency of trials in Blocks 3 and 4 to trials in Blocks 6 and 7. The accuracy of any given trial was ignored, and extreme latencies were recoded such that those less than 300 ms were scored as 300 ms and those greater than 3000 ms were scored as 3000 ms.³ The mean response latency for the prejudice-consistent block trials was subtracted from the mean response latency for the prejudice-inconsistent block trials. Thus, larger positive IAT effect scores reflected relatively stronger negative Black attitudes and relatively stronger positive White attitudes.

Explicit measures of prejudice. The five semantic differential scales revealed good reliability for Blacks ($\alpha = .91$) and for Whites ($\alpha = .89$). Thus, the mean of each set of scales was calculated, and a difference score was

² Readers may contact the authors for details about the behavior coding protocols.

³ Analyses were also conducted discarding responses in Blocks 3 and 6 (which presumably are more sensitive to task learning effects), as reported by Greenwald et al. (1998). Identical results obtained. Additional analyses using other trimming criteria (e.g., omitting incorrect trials, omitting trials with responses slower than 2 standard deviations from the mean) produced equivalent results.