Mere acceptance produces apparent attitude in the Implicit Association Test

Chris J. Mitchell[*]

School of Psychology, University of New South Wales, Sydney NSW 2052, Australia

Received 30 January 2003; revised 23 June 2003

Abstract

Two experiments examined mere acceptance effects in the Implicit Association Test (IAT). They tested whether accepting a stimulus as conforming to a rule produces responding consistent with positive attitude in the IAT. In Experiment 1, accepted stimuli were more easily categorized with pleasant personality characteristics than rejected stimuli; they were preferred according to the logic of the IAT. Accepted word stimuli were also responded to faster overall, suggesting that it was easier to make the accept than the reject response. In Experiment 2, numerical stimuli that conformed to a rule showed the same IAT preference effect over non-confirming stimuli, even when the rule conforming stimuli were more difficult to categorize. Three sources of this apparent preference for rule-conforming stimuli are considered: (1) the semantic relatedness of the concepts "accept" and "pleasant" on the one hand, and "reject" and "unpleasant" on the other; (2) that rejected non-category members are more salient ('pop-out') and thus are more easily categorized with the more salient unpleasant personality characteristics; or (3) that accepting rule-conforming stimuli is experienced as a pleasant event. Regardless of the mechanism underlying the mere acceptance effect, the IAT can produce apparent preferences for stimuli towards which participants have no positive attitude.

Attitudes are assumed to result from associative learning; "An *attitude* is the association of a social object or social group concept with a valence attribute concept" (Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellott, 2002; italics in original). Thus, a positive attitude towards a particular group results from an associative link between that group and attributes of positive valence such as generosity, warmth, and kindness. People are often found to associate positive attributes more with their own race than with other races. However, because such racist views are widely considered unacceptable, people often engage in conscious strategies to conceal them (Greenwald et al., 2002a, 2002b). To combat these strategies, researchers in social cognition have developed indirect, or "implicit," measures in the hope of detecting hidden racist views. The two most widely used procedures are affective priming (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and the Implicit Association Test (IAT; Greenwald,

McGhee, & Schwartz, 1998). This article focuses on the second of these procedures.

In the IAT, two categories of target items (e.g., insects and flowers) are categorized in two tests with two additional categories of attribute items that are known to be of positive and negative valence (e.g., pleasant and unpleasant personality characteristics). For example, in one test, a left key-press response is assigned to flowers and pleasant characteristics, and a right key press to insects and unpleasant characteristics. In this case, the assignment is affectively congruent; the two pleasant stimulus sets are categorized together, as are the two unpleasant stimulus sets. In another test, the response assignment to insects and flowers is reversed relative to the personality characteristics, and is thus affectively incongruent. A preference for flowers over insects is inferred if the "congruent" task is easier (responses are faster and more accurate) than the "incongruent" task. Test order is counterbalanced. Evidence for a wide range of implicit attitudes has been sought using the IAT including racial prejudice (e.g., Rudman, Greenwald, Mellott, & Schwartz, 1999), attitudes towards consumer durables

[*] Fax: +61-2-9385-3641.
E-mail address: cmitchell@psy.unsw.edu.au.

(Brunel, Collins, Greenwald, & Tietje, 1999) and self esteem (Greenwald & Farnham, 2000).

More recently, the IAT has been used to detect some very subtle attitudes, not based on well learned affective associations. For instance, in a study by Greenwald, Pickrell, & Farnham (2002b), participants studied four names of members of a hypothetical group for 45 s. When these names were compared with a second set of names in the IAT, there was an apparent preference for the studied names. The finding was interpreted as "implicit partisanship." That is, participants identified with the group comprising the four studied names and developed a preference for that group, but had no conscious awareness of the reason for this change in attitude (see Ashburn-Nardo, Voils, & Monteith, 2001 for a related finding).

It may be the case, as Greenwald et al. (2002a, 2002b), argue, that 45 s of study of four names is sufficient to induce a positive attitude toward these names. However, as Greenwald et al. themselves concede, there is a strong alternative explanation. The categorization process itself may have generated the apparent preference for the category members, a preference which was entirely unrelated to the perceived pleasantness of the target stimuli. Participants may have interpreted the IAT task as one of identifying, or "looking for" the studied names. These names would then have been treated as "correct" stimuli in the categorization task, and unstudied names as "incorrect." To explain the IAT effect in these terms, all that is required is a quite intuitive assumption that the concepts "correct" and "nice" on the one hand, and "incorrect" and "nasty" on the other hand, are semantically related. The semantic relatedness of pleasant words to the acceptance of the names as category members ('correct' stimuli) would predict the apparent preference for studied names, a preference unrelated to the participants' actual attitude towards the target items themselves.

A slightly different non-affective account of the IAT effect, which may also apply to Greenwald et al.'s data, was proposed by Rothermund and Wentura (2002). They suggested that, under some circumstances, the IAT may be treated as a salience matching task rather than an affective task. Thus, all unpleasant stimuli (e.g., insects and nasty personality characteristics) share the feature of high salience, and therefore are easily categorized together. That is, the unpleasant stimuli are the figure against the ground of the pleasant stimuli. Importantly, for the present analysis, the relative salience of items may be determined partly by the nature of the task. Thus, if participants are told (or assume) that a subset of the items are more or less important (perhaps by being given 45 s to study those items), then the IAT effect may reflect a processing bias based on the salience of the stimuli within the context of the task, not any intrinsic property of the stimuli themselves.

Greenwald et al. (2002b) acknowledged that their findings may reflect a mere categorization effect, but

they preferred implicit partisanship as an explanation. They argued that instructing the participants that the two sets of names were from two teams, and to imagine being a member of one teams, created implicit partisanship and thus the IAT effect. Greenwald et al. (2002b) argued that the evidence for a mere categorization effect in their data was, "at best suggestive and circumstantial." In general, if stimuli of differing affective value are compared in the IAT (however, subtle that difference may be), any effect observed may be due that difference and cannot unequivocally be attributed to a mere categorization effect. The only way to demonstrate a mere categorization effect of the kind Greenwald et al. describe is to compare two categories of stimuli of objectively identical affective value.

The present experiments used a similar structure to that of Greenwald et al. (2002b) to test whether mere categorization might be a sufficient explanation of their data. Two groups of stimuli of identical affective valence were compared in the IAT. One group of stimuli was "accepted" as conforming to a category rule, while the remaining stimuli were "rejected." It was expected that the items that conformed to the rule would be treated as the focal group (similar to the studied items in Greenwald et al.'s experiments) and would be more easily categorized with pleasant than unpleasant personality characteristics relative to the rejected stimuli (or preferred according to the logic of the IAT). This effect will be referred to as a "mere acceptance effect" rather than a "mere categorization effect," because only the former terminology indicates the specific direction of the IAT effect; a preference for the rule-conforming stimuli.

Experiment 1

To test the notion that the categorization process alone produces an IAT effect, two sets of stimuli were compared: animals or objects that possess teeth (Teeth stimuli) and animals or objects that fly (Flight stimuli). Flight stimuli did not posses teeth, and Teeth stimuli did not fly. Half of the participants were asked to categorize the stimuli as those with teeth versus those without (Teeth versus No Teeth). The remaining participants categorized the stimuli as those that fly versus those that do not (Flight versus No Flight). Thus, across participants, stimuli were equally often accepted as obeying the rule and rejected as not obeying the rule. It was predicted that the accepted stimuli would be preferred to the identical but rejected stimuli.

Method

Participants

Participants were 20 undergraduate students from the University of New South Wales, 14 females and 6 males,