Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Experimental Social Psychology 40 (2004) 357–365
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels☆

Cassandra L. Govan* and Kipling D. Williams*

Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia

Received 9 August 2002; revised 6 July 2003

Abstract

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a computer-based categorization task that measures concept association strengths. Greenwald et al. (1998) demonstrated that participants completed the categorizations more quickly when *pleasant* and *flower* shared a response key than when *pleasant* and *insect* shared a response key, and when *pleasant* and *White* shared a response key than when *pleasant* and *Black* shared a response key. In Study 1, we reversed the typical IAT effect for flowers and insects, and eliminated the typical IAT effect for White and Black, by changing the affective valence of the stimulus items. In Study 2, we replicated the reversibility effect for an animal and plant IAT, and supported a category re-definition hypothesis. Our results have implications for understanding the IAT, and suggest that the IAT not only measures stereotypic responses, but can also be influenced by individuating information of the stimulus items.

Keywords: Implicit Association Test (IAT); Stereotyping and individuating; Implicit attitudes; Category IAT

Imagine answering a multiple-choice question that asked, "what is your favorite animal?" The choices are dog, cat, bird, and fish. After answering this explicit measure, it is likely that, in addition to having assessed your attitude toward pets, your thoughts of animals are, at least temporarily, confined to animals that are pets. Now imagine the same question with alternatives of lion, tiger, elephant, and hippopotamus. Again, in addition to indicating an explicit attitude, the category "animal" is likely to be temporarily re-defined (or confined) to the subcategory jungle animal. In this paper, we argue that a similar process can occur in measuring responses on implicit measures of attitudes, like the Implicit Association Test (IAT; Greenwald et al., 1998).

Interest in implicit attitudes has led to a proliferation of measures that may bypass social desirability and conscious awareness. These measures include the IAT (Greenwald et al., 1998), Stereotypic Explanatory Bias (Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), a stereotype-primed lexical decision task (Wittenbrink, Judd, & Park, 1997), the sequential priming task (Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the Linguistic Intergroup Bias (von Hippel, Sekaquaptewa, & Vargas, 1997), the Go/No Go Association Task (Nosek & Banaji, 2001), and the Affective Simon task (De Houwer & Eelen, 1998), to name a few (for a review, see Fazio & Olson, 2003).

Probably the most widely used measure is the IAT (see Special Section in *Journal of Personality and Social Psychology: Attitudes and Social Cognition, Volume 81, 2001*, and the Special Issue of *Zeitschrift Für Experimentelle Psychologie, Volume 48, 2001*). In their original studies, Greenwald et al. (1998) found that when performing a categorization task in which two category labels are shared by one response key, and the other two category labels are shared by a different response key, participants were faster at the categorizations when *flower* and *pleasant* shared a response key than when *insect* and *pleasant* shared a response key (Study 1), and

☆ We thank Bill von Hippel, Anthony Greenwald, Jan De Houwer, and two anonymous reviewers for their comments on earlier drafts of this paper. The work reported in this paper represents part of a doctoral dissertation by Cassandra Govan, under the supervision of Kipling Williams.

* Corresponding authors.
E-mail addresses: cgovan@psy.mq.edu.au (C.L. Govan), kip@psy.mq.edu.au (K.D. Williams).

0022-1031/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2003.07.002

when *White* and *pleasant* shared a response key than when *Black* and *pleasant* shared a response key (Study 3).

Subsequent studies have provided support for these findings, with theoretical implications for racial stereotypes (Dasgupta, McGhee, Greenwald, & Banaji, 2000; De Houwer, 2001; McConnell & Leibold, 2001), self-esteem (Greenwald & Farnham, 2000), age stereotypes (Mellott & Greenwald, 1998), academic preferences (Nosek, Banaji, & Greenwald, 2002), and consumer preference and involvement (Maison, Greenwald, & Bruin, 2001; Williams, Govan, Edwardson, & Wheeler, 2001).

Some efforts have been aimed at determining how the IAT works and what aspects are most essential. In De Houwer's (2001) study, participants completed an IAT in which the category *British* was represented by three positive and three negative stimulus items, and the category *Foreign* was represented by three positive and three negative stimulus items. The results of De Houwer's (2001) study suggest that the IAT effect is driven by attitudes toward the category labels (e.g., British, Foreign), rather than attitudes toward the stimulus items (e.g., Princess Diana, Margaret Thatcher).

The proposition that stimulus items are relatively unimportant, however, has not been thoroughly scrutinized. Although there have been studies that suggest stimulus items may be important (e.g., Mitchell, Nosek, & Banaji, in press; Steffens & Plewe, 2001), we felt that a more thorough investigation of this issue was warranted.

Consider the standard insect/flower IAT, which arouses little attention or controversy. The stimulus items used for the flower category typically include items like rose and tulip. These items are not only flowers, but also pleasant exemplars of the flower category, thus confounding the category of flower with the category of pleasant. Similarly, the stimulus items used for the insect category typically include items like bee and wasp. These items are not only insects, but also unpleasant exemplars of the insect category, thus confounding the category of insect with the category of unpleasant. What would happen if this pattern of stimulus selection were reversed? That is, suppose the flower category contained rather unpleasant (albeit perhaps atypical) stimulus items like Venus flytrap and nettles? And suppose the insect category contained relatively pleasant (and perhaps atypical) items like butterfly and firefly? Will the category label still drive a flower–pleasant association, or as we predict, will the stimulus items influence the IAT effect? And, instead of using Theo (under the category Black) and Chip (under the category White), what would happen if we used Bill Cosby and Adolph Hitler?

Our procedure, then, is distinct from De Houwer's in that we saturate the stimulus array with all positively or all negatively valenced items, whereas De Houwer used half positive and half negative stimuli within the same array. Our task is fundamentally different from De Houwer's in that in our method, a single evaluative dimension can be associated with the entire stimulus array, as is the case with most IATs. Study 1 has two subcomponents, one dealing with flowers and insects (1a) and the other with Black and White (1b). Although they use the same participants (in which task is counterbalanced for which no order effects emerged), for ease of exposition we will report and analyze them in sequence.

Study 1a

Method

Participants and design

Eighty introductory psychology undergraduates (20 male, 60 female; M *age* $= 21.12$, $SD = 3.58$) were randomly assigned to complete either the *typical* IAT (in which stimulus items were positively valenced flowers and negatively valenced insects), or the *atypical* IAT (in which stimulus items were negatively valenced flowers and positively valenced insects).

Materials and procedure

The IAT was programmed using the Farnham Implicit Association Test (FIAT; Farnham, 1998). For both the typical and the atypical IATs, the stimulus items for *pleasant* were love, peace, happy, laughter, and pleasure, and the stimulus items for *unpleasant* were death, sickness, hatred, evil, and agony. For the typical IAT, the stimulus items for *flower* were rose, daffodil, daisy, violet, and poppy, and the stimulus items for *insect* were caterpillar, flea, cockroach, wasp, and maggot. For the atypical IAT, the stimulus items for *flower* were nettles, skunkweed, Venus flytrap, poison ivy, and weed, and the stimulus items for *insect* were ladybird,[1] butterfly, grasshopper, cricket, and firefly.

The IAT followed the standard blocks of categorization trials outlined by Greenwald et al. (1998). Block 1 consisted of 20 pleasant/unpleasant trials; Block 2 consisted of 20 flower/insect trials; Block 3 was a combined practice block of 20 trials; Block 4 was a combined data-collection block of 40 trials (the same label positions as practice Block 3); Block 5 consisted of 20 flower/insect trials (with labels in the reverse position of Block 2); Block 6 was a combined practice block of 20 trials (representing the new positions of flower/insect), and Block 7 was a combined data-collection block of 40 trials (the same label positions as practice Block 6).

[1] In Australia, ladybugs are called ladybirds.

Order was counterbalanced such that half the participants completed an IAT with *pleasant* and *insect* sharing a key in the first combined block, and half the participants completed an IAT with *pleasant* and *flower* sharing a key in the first combined block.

Participants completed the IAT in individual cubicles where they also received instructions and completed consent forms. After completion of the IAT, participants were fully debriefed, and thanked.

Results

Data reduction

As suggested by Greenwald et al. (1998), the first two trials of each data-collection block for all experiments were excluded from analysis, and reaction times that were shorter than 300 ms or longer than 3000 ms were re-coded to 300 and 3000 ms, respectively. Each participant's median reaction time for each data-collection block was calculated, and these medians were averaged to generate the group means for analysis.[2] No participants were excluded because of unusually high error rates.

IAT results

The counterbalancing factor of which combined block participants completed first did not influence the direction of the IAT effect; hence, further analyses are collapsed across this factor.

As shown in Fig. 1, participants who completed the typical stimulus item IAT responded faster when *pleasant* and *flower* shared a response key ($M = 686.70$, $SD = 146.75$), than when *pleasant* and *insect* shared a response key[3] ($M = 994.68$, $SD = 293.41$), $t(39) = -8.36$, and $p < .001$.

However, participants who completed the atypical stimulus item IAT responded faster when *pleasant* and *insect* shared a response key ($M = 794.60$, $SD = 204.39$), than when *pleasant* and *flower* shared a response key ($M = 870.88$, $SD = 202.22$), $t(39) = 2.14$, and $p = .039$.

Examination of the difference scores (the difference between the *pleasant* and *insect* reaction time and the *pleasant* and *flower* reaction time) reveals that the IAT effect for the typical condition ($M = 307.98$, $SD = 233.05$) was significantly larger than the IAT effect for the atypical condition ($M = -76.28$, $SD = 225.41$), $t(78) = 4.52$, and $p < .001$ (of interest here is the absolute value of the difference score, not the sign, which indicates direction of the effect).

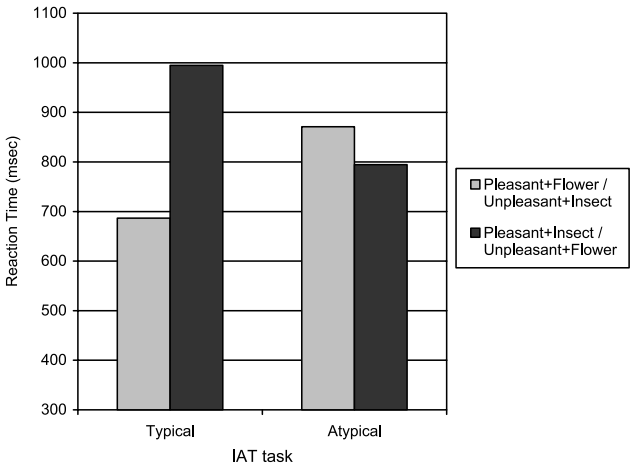

Fig. 1. IAT results for Study 1a. Reaction times as a function of IAT task (typical or atypical) and categorization pairing (pleasant + flower/unpleasant + insect, or pleasant + insect/unpleasant + flower).

Discussion

Our aim in Study 1a was to investigate whether category labels solely drove the IAT effects. We manipulated the valence of flower and insect stimulus items under their respective category labels. The results from our typical IAT support the results of Greenwald et al. (1998). However, our atypical IAT revealed a reversal, leading to the rather implausible inference of a preference for insects over flowers. This preference, however, was not as large as we found for the typical IAT, which may be related to the atypicality of the stimulus items. Our results, therefore, suggest that stimulus items can influence the IAT effect.

In Study 1b, we examine whether we can reverse the typical IAT effect for the Black/White IAT. As in Study 1a, we predict a reversal of the typical IAT effect when the affective valence of the exemplars is saturated with pleasant Black names and unpleasant White names.

Study 1b

Whereas no published studies have un-confounded (or counter-confounded) the valence of the flower and insect categories with the pleasant and unpleasant categories, there have been a few published studies that bear some conceptual similarity to this process for race or country of origin.

De Houwer (2001) examined target concept trial reaction times with a category set that included both typical and atypical stimulus items. He found that "responses to positive British and negative foreign names were no faster than responses to negative British and positive foreign names" (De Houwer, 2001, p. 448). However, we are more interested in what happens to the

[2] Throughout this paper, we report analyses of the means of medians, but in all cases, similar results were obtained for means of medians, means, and log-transformed means.

[3] When *pleasant* and *flower* shared a response key, *unpleasant* and *insect* shared the alternative response key. For ease of reading, we will use this abbreviated reporting throughout this paper.

overall IAT effect (i.e., a closer association between *White* and *pleasant* or between *Black* and *pleasant*) when the stimulus items for Black and White are *all* selected to affectively favor Blacks.

In Dasgupta and Greenwald's (2001) study, participants completed an ostensible general knowledge test in which they had to identify a number of famous individuals, followed by a standard Black/White IAT. The individuals to be identified in the general knowledge test were either admired Black and disliked White individuals, or disliked Black and admired White individuals. The results showed that the pro-White IAT effect was substantially reduced (but not eliminated) by exposing participants to disliked White and admired Black exemplars prior to the completion of the IAT. Our Study 1b will examine the impact on IAT effects when similarly admired and disliked exemplars comprise the stimulus items in the IAT.

In research conducted concurrently with ours, Mitchell et al. (in press, Study 2) found that a Black/White IAT consisting of disliked Black and liked White exemplars produced the typical pro-White IAT effect; whereas a stimulus array consisting of liked Black and disliked White exemplars yielded no significant IAT effect. We use a different comparison: rather than using disliked Black and disliked White exemplars, we examine the difference between a standard Black/White IAT and a disliked White/liked Black IAT.

Method

Participants and design

Participants in Study 1b were the same participants as in Study 1a. Counterbalancing was used such that participants either completed a flower/insect IAT or a Black/White IAT first, and they did one typical IAT and one atypical IAT (order counterbalancing did not influence the direction of IAT effects).

Materials and procedure

All procedures used in Study 1b were the same as Study 1a, including the category labels *pleasant* and *unpleasant*, and their respective stimulus items. For the typical IAT, the stimulus items for *Black* were Theo, Leroy, Tyrone, Lakisha, and Ebony, and the stimulus items for *White* were Chip, Josh, Todd, Amber, and Betsy. For the atypical IAT, the stimulus items for *Black* were Michael Jordan, Bill Cosby, Eddie Murphy, Cathy Freeman, and Ernie Dingo, and the stimulus items for *White* were Charles Manson, Adolph Hitler, Hannibal Lechter, Pauline Hanson, and Martin Bryant.[4]

[4] Australians know Pauline Hanson as a disliked politician, Martin Bryant as a mass-murderer, Cathy Freeman as a popular athlete, and Ernie Dingo as a popular actor.

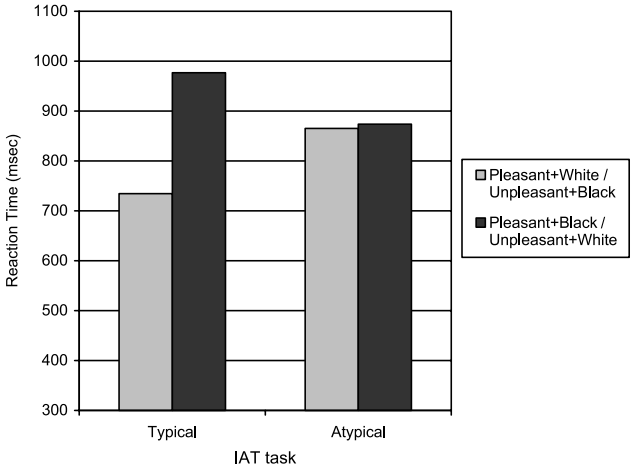

Fig. 2. IAT results for Study 1b. Reaction times as a function of IAT task (typical or atypical) and categorization pairing (pleasant + white/unpleasant + black, or pleasant + black/unpleasant + white).

Results

IAT results

Data reduction techniques were the same as outlined in Study 1a. As in Study 1a, the counterbalancing factor of which combined block participants completed first did not influence the direction of the IAT effect; hence, further analyses are collapsed across this factor.

As shown in Fig. 2, and consistent with Greenwald et al. (1998), participants who completed the typical IAT responded faster when *pleasant* and *White* shared a response key ($M = 734.39$, $SD = 210.66$), than when *pleasant* and *Black* shared a response key ($M = 976.63$, $SD = 192.87$), $t(39) = -7.17$, and $p < .001$.

However, participants who completed the atypical IAT were no faster when *pleasant* and *White* shared a response key ($M = 865.09$, $SD = 170.30$), than when *pleasant* and *Black* shared a response key ($M = 873.65$, $SD = 222.38$), $t(39) = -0.22$, and $p = .82$.

The difference score for the typical condition ($M = 242.24$, $SD = 213.55$) was significantly larger than the atypical condition ($M = 8.56, SD = 241.57$), $t(78) = 4.58$, and $p < .001$.

Discussion

In this study, we replicated previous pro-White IAT effects when the stimulus items were commonly recognized first names of White and Black individuals. The comparable magnitude of this replication as that found by Greenwald et al. (1998, Study 3) is noteworthy because our participants were Australian, for whom the typical American first name stimulus items may have been less familiar. However, when we used positively valenced famous Black names and negatively valenced famous White names, the pro-White IAT effect disappeared.

Our speculation as to why we did not obtain a reversal is that there is still some element of a stronger association between *White* and *pleasant*, even when the White exemplars are undesirable, or even evil. Perhaps own-race preference, compared to flower preference, is a more practiced and culturally instilled attitude. We propose, and examine further in Study 2, that stimulus items lead to category re-definition. Perhaps own-race preference is more resistant to this category re-definition.

The finding of an elimination, but no reversal, of the pro-White IAT effect in the atypical IAT is consistent with Dasgupta and Greenwald's (2001) finding of reduced pro-White IAT effect following exposure to atypical exemplars. This suggests an alternative explanation to their findings, that pre-exposure to admired Blacks and undesired Whites temporarily re-defined the categories of Blacks and Whites. Our results are also consistent with those obtained by Mitchell et al. (in press). In Study 2, we examine the process behind the stimulus item influence.

Both Studies 1a and 1b demonstrated that the stimulus items can influence the IAT effect, and therefore the selection of stimulus items should be considered carefully. Although we infer that participants are temporarily re-defining the categories as a function of the stimulus items, we aim to support this proposition more directly in Study 2.

Study 2

To gain an understanding of the processes involved in our atypical IATs, and in IATs in general, we conducted a second study, in which all participants completed an animal/plant, pleasant/unpleasant IAT (Stage 1). We chose animal/plant to more correctly define the stimulus items, and to allow greater ease in selecting positively and negatively valenced exemplars of each category. Half the participants completed an IAT for which the stimulus items were comprised of negatively valenced (i.e., *nasty*) animals and positively valenced (i.e., *nice*) plants, and half the participants completed an IAT in which the stimulus lists were comprised of *nice* animals and *nasty* plants.

Following the IAT, all participants completed a filler task (Stage 2; a lexical decision task), and then completed what we called a *Category IAT* (Stage 3). In the Category IAT, the category labels were pleasant/unpleasant and animal/plant. The stimulus items for *pleasant* and *unpleasant* were the same as in the Stage 1 IAT. To remove the influence of stimulus items on IAT effects, the stimulus items for the categories *animal* and *plant* were simply the words "animal" and "plant."

Our hypothesis was that if participants were re-defining the category *animal* to *nice animal* (in Stage 1), then they should be faster at the Category IAT when *pleasant* and *animal* shared a response key than when *pleasant* and *plant* shared a response key. Similarly, if participants were re-defining the category *plant* to *nice plant* (in Stage 1), then they should be faster at the Category IAT when *pleasant* and *plant* shared a response key than when *pleasant* and *animal* shared a response key.

Method

Participants and design

Participants were 67 undergraduate psychology students (16 male, 51 female), ranging in age from 18 to 53 years ($M = 21.05$, $SD = 6.59$), who received credit for participation. Thirty-two participants completed the nasty animal/nice plant Stage 1 IAT, and 35 participants completed the nice animal/nasty plant Stage 1 IAT. All participants then completed the filler task followed by the Category IAT.

Materials and procedure

All tasks were programmed using DirectRT and MediaLab (Jarvis, 2002a, 2002b). For both versions of the IAT in Stage 1, the stimulus items for *pleasant* were sunrise, smile, joy, happy, and peace, and the stimulus items for *unpleasant* were vomit, war, hate, agony, and death. For the nasty animal/nice plant IAT, the stimulus items for *animal* were crocodile, grizzly bear, black snake, maggot, and pit-bull, and the stimulus items for *plant* were daffodil, lily, chrysanthemum, carnation, and daisy. For the nice animal/nasty plant IAT, the stimulus items for *animal* were seahorse, swan, puppy, joey, and bunny rabbit, and the stimulus items for *plant* were poison ivy, pondweed, Venus flytrap, thornbush, and sword grass.

The IAT followed the same ordering of blocks as described in Studies 1a and 1b. We dropped the counterbalancing factor of which combined block participants completed first, and instead all participants completed the congruent block first.

At Stage 2, participants completed a filler task. Participants were presented with a letter string in the center of the computer screen, and their task was to decide whether the letter string was a word or a non-word. The words consisted of 10 positively valenced plants, 10 positively valenced animals, 10 negatively valenced plants, and 10 negatively valenced animals (these items differed from the ones used in the Stage 1 IAT). The non-word list was compiled using an online non-word generator (http://www.maccs.mq.edu.au/~nwdb/) and were matched in length to the words. Each letter string was randomly presented twice.

In pilot studies, we found this lexical decision task to be unaffected by IAT conditions (see Becker, Moscovitch, Behrmann, & Joordens, 1997; Joordens & Becker,

1997, for arguments against the sensitivity of lexical decision tasks). We saw it as a useful and unbiased filler task because we exposed the participants to the whole range of plants and animals.

For the Stage 3 IAT, the stimulus items for *pleasant* and *unpleasant* were the same as the Stage 1 IAT. The stimulus item for *animal* was "animal," and the stimulus item for *plant* was "plant." The order of which block participants completed first (*pleasant* and *plant* sharing a response key first, or *pleasant* and *animal* sharing a response key first) was counterbalanced.

Participants completed the experiment in individual cubicles where they also received instructions and completed consent forms. Once finished, participants were fully debriefed, and thanked.

Results

Data reduction techniques for the IATs were the same as outlined in Studies 1a and 1b. There were no significant differences on the Stage 2 lexical decision filler task, thus these results will not be discussed further.

Stage 1 IAT results

As shown in Fig. 3, participants who completed the nice animal/nasty plant IAT, responded faster when *pleasant* and *animal* shared a response key ($M = 643.01$, $SD = 84.20$), than when *pleasant* and *plant* shared a response key ($M = 887.07$, $SD = 162.82$), $t(34) = 10.81$, and $p < .001$.

Participants who completed the nasty animal/nice plant IAT responded faster when *pleasant* and *plant* shared a response key ($M = 605.44$, $SD = 84.93$), than when *pleasant* and *animal* shared a response key ($M = 905.50$, $SD = 199.06$), $t(31) = 10.28$, and $p < .001$.

Stage 3 IAT results

As shown in Fig. 4, participants who completed the nice animal/nasty plant IAT in Stage 1 responded faster in the Category IAT when *pleasant* and *animal* shared a response key ($M = 650.04$, $SD = 80.83$), than when *pleasant* and *plant* shared a response key ($M = 725.47$, $SD = 139.59$), $t(34) = 3.74$, and $p = .001$.

Participants who completed the nasty animal/nice plant IAT in Stage 1 responded faster in the Category IAT when *pleasant* and *plant* shared a response key ($M = 660.72$, $SD = 112.93$), than when *pleasant* and *animal* shared a response key ($M = 765.02$, $SD = 147.01$), $t(31) = -5.09$, and $p < .001$.

Discussion

In Stage 1, the valence of the stimulus items influenced the IAT effects. Participants in the nice animal/nasty plant IAT condition responded faster when *pleasant* and *animal* shared a response key than when *pleasant* and *plant* shared a response key. Participants in the nasty animal/nice plant IAT condition responded faster when *pleasant* and *plant* shared a response key than when *pleasant* and *animal* shared a response key.

The addition of a Category IAT at Stage 3 gave support to the re-definition of categories hypothesis. Results in the Stage 3 IAT were influenced by the Stage 1 IAT condition, such that participants who were presented with nice animals and nasty plants in the Stage 1 IAT were faster at the Category IAT when *pleasant* and *animal* shared a response key than when *pleasant* and *plant* shared a response key, and the opposite pattern emerged for participants who were presented with nasty animals and nice plants in the Stage 1 IAT.

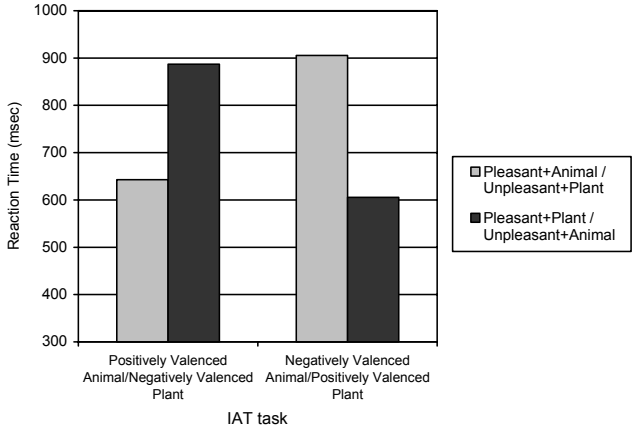

Fig. 3. Results for Stage 1 IAT in Study 2. Reaction times as a function of IAT task (positively valenced animal/negatively valenced plant or negatively valenced animal/positively valenced plant) and categorization pairing (pleasant + animal/unpleasant + plant, or pleasant + plant/unpleasant + animal).

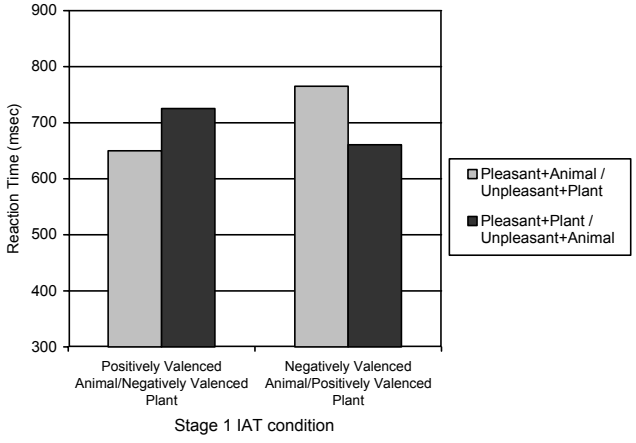

Fig. 4. Results for Stage 3 IAT in Study 2. Reaction times as a function of Stage 1 IAT condition (positively valenced animal/negatively valenced plant or negatively valenced animal/positively valenced plant) and categorization pairing (pleasant + animal/unpleasant + plant, or pleasant + plant/unpleasant + animal).

General discussion

Both studies provide support for the hypothesis that IAT effects are not solely a function of category labels. The stimulus items chosen to represent the category labels are also important, and they may drive participants to re-define the categories. Rather than being merely a methodological footnote for IAT research, we think there are at least two ways to interpret these results at a broader level.

Our findings may reveal something more fundamental about implicit attitudes. Perhaps these results speak to individuals' abilities to disregard category labels and stereotypes if they are sufficiently infused with members of a category that are incongruent with their stereotypes. Our results in Study 1b, showing an elimination of a pro-White IAT effect, are inconsistent with stereotype research showing that non-stereotyped responses require greater cognitive effort. Devine (1989) demonstrated that when participants could not consciously monitor their responses, both high and low prejudiced individuals gave responses consistent with stereotypes. Furthermore, studies examining responses when participants are cognitively busy, responding at a non-optimal time of day, or angry, also show an increased reliance on stereotypes (e.g., Bodenhausen, 1990; Macrae, Milne, & Bodenhausen, 1994; Rogers & Prentice-Dunn, 1981), even if participants are motivated to respond non-stereotypically (Pendry & Macrae, 1994; but for further boundary conditions for these effects see Fein & Spencer, 1997; Gilbert & Hixon, 1991). It should also be noted that Fazio et al. (1995) found that some individuals did not show automatic activation of prejudiced responses. Thus, despite some exceptions, the literature suggests that to respond without the use of stereotypes requires conscious effort, and that automatic responding should be stereotype-based. Accordingly, IAT results should show the stereotyped response, regardless of the stimulus items used, because the task requires responses that are too fast to consciously monitor and alter. Nevertheless, our flower/insect atypical IAT showed a reversal of the pro-flower effect, and our Black/White atypical IAT showed an elimination of the pro-White effect, suggesting that participants were able to over-ride the stereotypical response automatically.

However, is it really plausible to suggest that eliminating the typical IAT effect in an IAT containing positive Black stimulus items and negative White stimulus items demonstrates an overriding of implicit stereotypes, or a change in the implicit attitudes towards these racial groups? Perhaps a more plausible interpretation of the results, still consistent with the re-definition of categories hypothesis, is that the IAT effects are influenced when participants can draw individuating information from the stimulus items. In the standard IAT, the categories Black and White are represented by first names that are typical of each race. Thus, participants have nothing else to latch on to except their stereotypical view of each race; hence, the IAT shows a stereotypic effect. However, when we infuse the stimulus array with items that either allow individuating information to be inferred (e.g., Adolph Hitler, Michael Jordan), or a subcategory to become activated (e.g., kitten, tiger), we see IAT effects that might represent something different from the general stereotypic response.

This second interpretation is reminiscent of research on stereotyping versus individuating (Fiske, Lin, & Neuberg, 1999; Fiske, Neuberg, Beattie, & Milberg, 1987). Although their studies used explicit responses in a person perception task, perhaps our studies demonstrate that similar processes may be at work in implicitly measured responses. The similarity is evident in their explanation of the stereotyping/individuating response: "If the category is judged to fit, then responses are relatively category-based. If not, then responses are relatively more individuating, or attribute-oriented" (Fiske et al., 1987, p. 403). We think this is an accurate way to describe what might happen in an IAT. If the stimulus set fits the category label, responses will be stereotypic. If the stimulus set does not fit the category label in a stereotypic way (but still consists of accurate exemplars of that category), responses will be individuating. Our extension on this idea is that if responses are individuating, then the category label may temporarily be re-defined to match the set of stimulus items (e.g., *nice animals*).

Our findings do not pose an intractable problem for the IAT. Rather, they provide us with information that we can use for pre-testing exemplars, interpretation of IAT results, and interpretation of what we are measuring with the IAT. The conception that the IAT is a measure of the *true* attitude (a view made more strongly by media reports than by the creators of the IAT, e.g., "Roots of Racism Revealed" http://www.abcnews.go.com/sections/living/InYourHead/allinyourhead_11.html) might need to be altered. At the very least, this conception should be altered to view the IAT as a useful way of implicitly measuring an attitude toward a concept, where that concept is defined by the representing set of stimulus items.

Our findings are also not incompatible with those of De Houwer (2001). Although De Houwer suggests that the category labels over-ride the stimulus items, the stimulus item lists in his study were mixed, and thus, his task may be a conceptually different task for the participants. Participants would not be able to re-define the category of *foreign* to *disliked foreign* when the list includes both liked and disliked foreign names. Furthermore, we are suggesting that stimulus items promote a re-definition of the category labels, and this new definition of the category label influences the IAT effect. If it were simply the stimulus items that drove the effect, we

would have found a significant pro-Black effect in Study 1b. If it were simply the category labels, we would have found a significant pro-White effect in Study 1b. Instead, what we found is the influence of the stimulus items on the category definition, which then determines the IAT effect. The category IAT of Study 2 provides direct evidence that category re-definition is a reasonable explanation for our effects. Thus, we see our studies as complementing the work done by De Houwer, and taken together, they may help in uncovering the processes underlying the IAT.

What does this mean for the IAT?

From our view, two broad reactions have surfaced in response to the IAT. On the one hand, there appears to be a rush to apply the IAT to everything, as though we understand fully its meaning and implications, and as if it is the Holy Grail of attitude and self-esteem measurement (see Glick, 2002). On the other hand, there has been a backlash against the IAT, in which it has been suggested that it is meaningless (see Brendl, Markman, & Messner, 2001; Fiedler, Messner, & Bluemke, 2003). Our position is somewhere in the middle. Our studies have shown that the stimulus items chosen can influence the IAT, and that the process underlying this might be category re-definition. Nevertheless, we also see the IAT as a useful tool, as long as these limitations are taken into consideration. If the goal of the research is to demonstrate the magnitude of an attitude, then the selection of stimulus items could be crucial. If, on the other hand, the IAT is used as a dependent measure that is hypothesized to show greater or lesser associations as a function of prior manipulations, then the stimulus items may be of less concern.

References

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A conceptual account and empirical evidence. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*, 1059–1082.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 1*, 319–322.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 760–773.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800–814.

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

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*, 443–451.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12*, 45–61.

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

Farnham, S. (1998). The Farnham Implicit Association Test for Windows, Version 2.3.

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.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.

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

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*, 31–44.

Fiedler, K., Messner, C., & Bluemke, M. (2003). *Applying psychometric criteria to attitude measurement using the implicit association test: A theoretical, empirical, and ethical appraisal*. University of Heidelberg, unpublished manuscript.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford Press.

Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology, 23*, 399–427.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Glick, P. (2002). Help Tony's Kids!: Signs and symptoms of IAT-OCD. *Dialogue, 17*, 12–13.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

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*, 1022–1038.

Jarvis, W. B. G. (2002a). MediaLab v2002. Computer Program for Windows.

Jarvis, W. B. G. (2002b). DirectRT v2002. Computer Program for Windows.

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*, 1083–1105.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 37–47.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 2*, 61–79.

McConnell, A. R., & Leibold, J. M. (2001). Relations between the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435–442.

Mellott, D. S., & Greenwald, A. G. (1998, April). *Do older adults show automatic ageism*? Poster session presented at the Annual Meeting of the Midwestern Psychological Association, Chicago, IL.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (in press). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625–666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math is not equal to me. *Journal of Personality and Social Psychology, 83*, 44–59.

Pendry, L. F., & Macrae, C. N. (1994). Stereotypes and mental life: The case of the motivated but thwarted tactician. *Journal of Experimental Social Psychology, 30*, 303–325.

Rogers, R. W., & Prentice-Dunn, S. (1981). Deindividuation and anger-mediated interracial aggression: Unmasking regressive racism. *Journal of Personality and Social Psychology, 41*, 63–73.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology, 39*, 75–82.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift fur Experimentelle Psychologie, 48*, 123–134.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33*, 490–509.

Williams, K. D., Govan, C. L., Edwardson, M., & Wheeler, L. (2001, February). Consumer involvement can be measured by the Implicit Association Test. *Conference Presentation at the Society of Consumer Psychology*, Scottsdale, Arizona.

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.