JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2007, **88,** 263–283 NUMBER 2 (SEPTEMBER)

A DERIVED TRANSFER OF FUNCTIONS AND THE IMPLICIT ASSOCIATION TEST

CATRIONA O'TOOLE, DERMOT BARNES-HOLMES, AND SINEAD SMYTH

NATIONAL UNIVERSITY OF IRELAND, MAYNOOTH

Participants were trained in a series of interrelated conditional discriminations that aimed to establish four 4-member equivalence classes (i.e., A1–B1–C1–D1, A2–B2–C2–D2, A3–B3–C3–D3, A4–B4–C4–D4). During this training, the four A stimuli (i.e., A1, A2, A3, and A4) were compounded with pictures containing positive or negative evaluative functions (A1/A2 negative & A3/A4 positive). The transfer of evaluative functions to directly and indirectly related members of the equivalence classes (i.e., B, C, and D stimuli) was measured using an Implicit Association Test (IAT). During consistent test blocks, participants were required to press the same response key for target words that were related to those A stimuli that possessed similar evaluative functions (A1/A2-left key & A3/A4-right key). During inconsistent test blocks, target words that were related to those A stimuli with different evaluative functions were assigned to the same response key (A1/A4–left key & A2/A3–right key). Results showed that all 8 participants, who passed a matching-to-sample equivalence test following the IAT, responded more rapidly on consistent relative to inconsistent test blocks. This typical IAT effect was not observed for those participants who did not pass the equivalence test. The results suggest that the IAT effect may arise from formally untested derived relations, and supports the argument that such relations could provide a valid behavioral model of semantic categories in natural language.

Key words: implicit association test, derived relations, transfer of functions, matching to sample, adults

---

The derived transfer of functions has been well documented in the behavior-analytic literature (e.g., Barnes & Keenan, 1993; Hayes, Kohlenberg, & Hayes, 1991; Roche & Barnes, 1997). A derived transfer occurs when the function of one or more members of an equivalence class transfers to other members in the absence of an explicit or direct training history. Consider a participant who is trained and tested for the formation of a three member equivalence class, A–B–C. If the A stimulus is then used to predict the delivery of a mild electric shock, that stimulus may well acquire an aversive function such that the participant may report fear and produce signs of physiological arousal in its presence. In addition, the other members of the equivalence class also may acquire aversive functions, although they have not been directly paired with shock (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994). In effect, the aversive function established for the A stimulus through direct stimulus pairing transfers to the equivalently related B and C stimuli. This latter effect is difficult to explain in terms of classical conditioning alone because neither the B nor C stimuli have been used to predict shock (see Smyth, Barnes-Holmes, & Forsyth, 2006, for a detailed discussion).

In many studies the methodology for testing derived transfer allows the participant to generate a self-rule that involves ''working out'' what the experimenter is looking for (e.g., Barnes & Keenan, 1993; Wulfert & Hayes, 1988), and then behaving accordingly. Although this issue may not raise conceptual or theoretical difficulties for certain research questions, in other areas it seems to threaten external validity. For example, if a derived transfer of functions is being used to develop a model of fear acquisition, it is important that a transfer of eliciting functions is recorded, and not just an operant avoidance response that could be controlled in part by the participants' desire to please the experimenter. A small number of studies have therefore employed psychophysiological measures in the

The preparation of this manuscript was funded in part by a scholarship from the Irish Research Council for the Humanities and Social Sciences and by the John and Pat Hume Scholarship from the National University of Ireland, Maynooth. The authors would like to thank Tony Greenwald for his regular advice and assistance concerning all aspects of the IAT. We also thank Jan De Houwer and Paul Smeets for their constructive comments on an earlier draft of the current article.

Sinead Smyth is now at the Department of Psychology, Royal Holloway, University of London.

Correspondence concerning this article should be addressed to Catriona O'Toole (catriona.a.otoole@nuim.ie) or to Dermot Barnes-Holmes (Dermot.Barnes-Holmes@nuim.ie), Department of Psychology, National University of Ireland, Maynooth, Maynooth, Co. Kildare, Ireland.

doi: 10.1901/jeab.2007.88-263

assessment of derived transfer effects (e.g., Dougher et al., 1994; Roche & Barnes, 1997) in order to demonstrate that the transferred function possessed the appropriate emotional or arousal functions.

Although the use of physiological measures provides one possible means of controlling for experimenter demand effects in assessing the derived transfer of functions, it would seem prudent to explore other methodologies. One relevant method that has been widely used in social and clinical psychology research is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT involves presenting two attribute concepts such as ''pleasant'' and ''unpleasant,'' and two target concepts such as ''insect'' and ''flower'' in the top right and left hand corners of a computer screen (see Figure 1). Target stimuli pertaining to these four concepts (e.g., ''love,'' ''abuse,'' ''spider,'' and ''tulip'') are then presented in the center of the screen and participants must respond to them by selecting the relevant response key. In one task, ''pleasant'' and ''flower'' are grouped together and share a single response key, as do ''unpleasant'' and ''insect.'' It is assumed that ''pleasant'' and ''flower'' are associated in memory, whereas ''insect'' is more likely to be associated with ''unpleasant,'' and thus these tasks are termed consistent category tasks and responding is predicted to be relatively fast. In a second task the response assignment for ''insect'' and ''flower'' is reversed whereas the response assignment for ''pleasant'' and ''unpleasant'' remains the same. Thus, the unrelated concepts, ''pleasant'' and ''insect,'' share one response key and ''unpleasant'' and ''flower'' share the other key. These tasks are referred to as inconsistent and responding is now predicted to be slower, relative to the consistent tasks. The difference in response times between consistent and inconsistent tasks is known as the IAT effect (Greenwald, Nosek, & Banaji, 2003).

A widely agreed technical explanation for the IAT effect is not yet available (e.g., De Houwer, 2001; Moors & De Houwer, 2006). In general terms, however, slower responding during the inconsistent tasks emerges because some form of response-competition is created by the incompatible valences of the stimuli. For example, it seems likely that most participants will respond emotionally to insects as unpleasant and flowers as pleasant. Critically, however, the inconsistent task requires responding against this tendency by categorizing insects with pleasant and flowers with unpleasant. This type of response competition is not present during consistent tasks, because like is categorized with like, and thus correct responding on average should be faster for consistent tasks (see De Houwer, 2003, for a detailed discussion and a review of the relevant evidence).

Mitchell, Anderson, and Lovibond (2003) were among the first to generate the IAT effect with laboratory-induced associations. In their study, participants were trained to pair nonsense stimuli with pleasant and unpleasant words using traditional classical conditioning procedures. The transfer of affect from the pleasant and unpleasant words to the nonsense stimuli was measured in the IAT. The results showed that nonsense stimuli given pleasant meanings in training were more rapidly categorized with pleasant than unpleasant personality characteristics, compared to nonsense stimuli given unpleasant meanings. Thus the study produced a difference in reaction times across consistent versus inconsistent tasks, thereby demonstrating that the IAT is sensitive to laboratory-induced associations. The results reported by Mitchell et al. (2003) indicate that the IAT effect can be produced with nonsense words when those words have been directly paired with affectively valenced stimuli, but it remains to be seen if a similar effect is observed for indirectly related stimuli using a derived transfer of function procedure.

The current study involved training and testing four equivalence relations, two of which had negative evaluative functions (spiders and snakes) and the other two having positive evaluative functions (babies and romance). These equivalence relations then were substituted for real words in an IAT to determine if they produced an IAT effect like that observed using words from natural language (see Table 1). The first part of the experiment involved training participants in 12 matching-to-sample (MTS) tasks that were designed to establish four 4-member equivalence classes: A1–B1–C1–D1, A2–B2–C2–D2, A3–B3–C3–D3, and A4–B4–C4–D4. In addition, the A1 stimulus was directly paired with pictures of spiders, A2 with pictures of snakes, A3 with pictures of babies, and A4 with

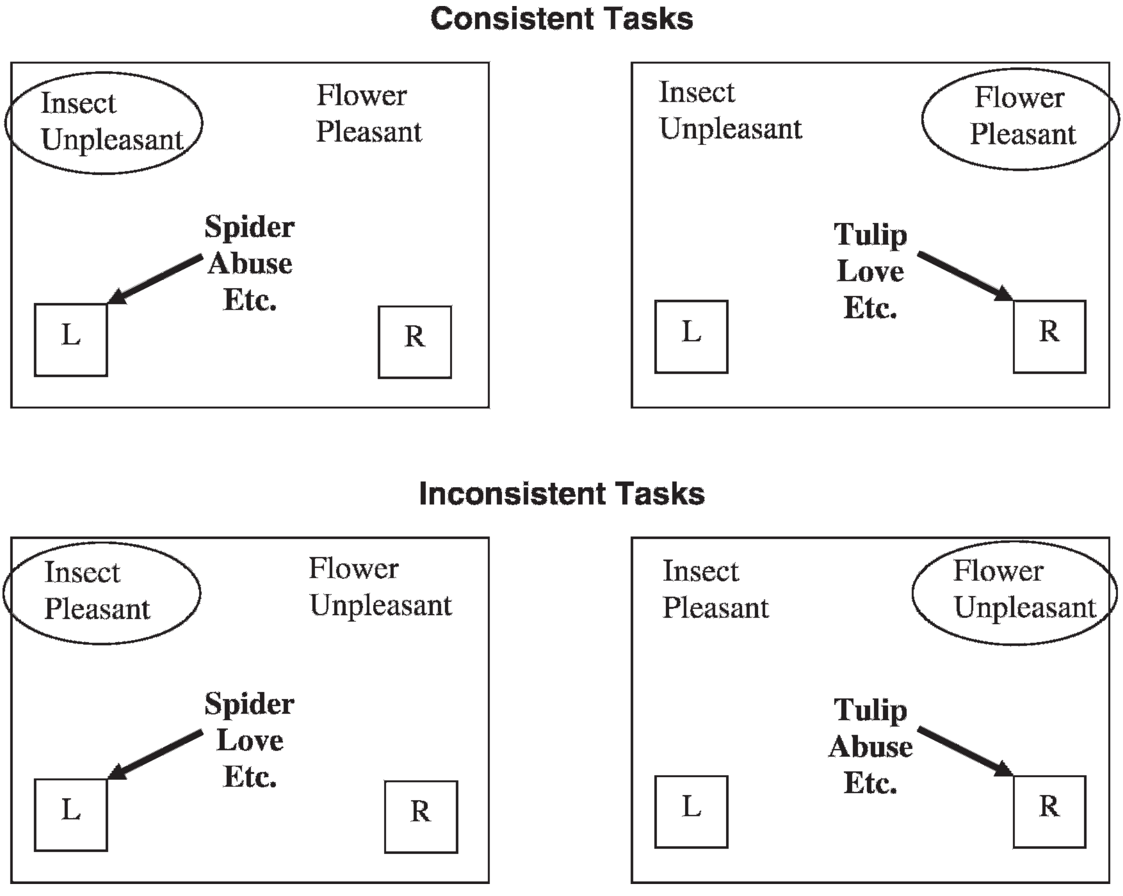

Fig. 1. Schematic representation of sample Implicit Association Test (IAT) trials. The two upper boxes show consistent trials, which involve categorizing stimuli together that are assumed to be strongly associated. The top-left box shows a trial in which the participant must press the left key when an insect or unpleasant target word is presented; the top-right box shows a trial for which the right key is pressed when a flower or pleasant target word is presented. The lower two boxes show inconsistent trials, which involve categorizing stimuli together that are assumed *not* to be strongly associated. The bottom-left box shows a trial in which the participant must press the left key when an insect or pleasant target word is presented; the bottom-right box shows a trial for which the right key is pressed when a flower or unpleasant target word is presented. (Note: the arrows and circles were not presented during the actual IAT procedure.)

pictures of romantic, heterosexual couples. If this training successfully generates a derived transfer of evaluative functions, all four stimuli in each of the four classes, 1, 2, 3, and 4, should acquire spider, snake, baby, and romance functions, respectively. For the next part of the experiment, it was predicted that participants should produce faster IAT responding when the response assignment for equivalence relations is consistent with the evaluative functions of the four classes (i.e., the spider and snake classes are assigned to one key and the baby and romance classes are assigned to the other key) than when the response assignment is inconsistent with such evaluative functions (e.g., snake and romance classes assigned to one key and spider and baby classes to the other key).

The current research was designed also to determine if stimulus pairings or associations per se were sufficient to generate the IAT effect or if equivalence class formation was required. To answer this question, all participants were trained to a criterion on the MTS tasks (thus establishing reliable stimulus associations), but were only exposed to an equivalence test following exposure to the IAT. If associations per se are sufficient to produce the IAT effect then it should be observed for all participants who successfully complete the MTS training. If,

Table 1

English words used for the baseline IAT and the two stimulus sets used for the MTS training/retraining, derived IAT, and equivalence test.

Stimuli used in the baseline IAT
Spider: Creep, Web, Tarantula
Snake: Slither, Hiss, Viper
Baby: Toddle, Pram, Milk
Romance: Dating, Roses, Engagement
Stimulus sets used with MTS and the derived IAT
Set 1:
A1 = JEP / Four pictures of spiders (Spider)*
B1 = VIR (Creep)
C1 = PUK (Web)
D1 = ROG (Tarantula)
A2 = POF / Four pictures of snakes (Snake)
B2 = BOC (Slither)
C2 = KEL (Hiss)
D2 = ZID (Viper)
A3 = TOB / Four pictures of Babies (Baby)
B3 = KED (Toddle)
C3 = YUB (Pram)
D3 = MAU (Milk)
A4 = MIQ / Four pictures of Romance (Romance)
B4 = VUD (Dating)
C4 = ZAT (Roses)
D4 = DAX (Engagement)
Set 2: Class 1 = Babies and Class 3 = Spiders; Classes 2 and 4 are the same as Set 1
A1 = JEP / Four pictures of Babies (Baby)
B1 = VIR (Toddle)
C1 = PUK (Pram)
D1 = ROG (Milk)
A2 = POF / Four pictures of snakes (Snake)
B2 = BOC (Slither)
C2 = KEL (Hiss)
D2 = ZID (Viper)
A3 = TOB / Four pictures of spiders (Spider)
B3 = KED (Creep)
C3 = YUB (Web)
D3 = MAU (Tarantula)
A4 = MIQ / Four pictures of Romance (Romance)
B4 = VUD (Dating)
C4 = ZAT (Roses)
D4 = DAX (Engagement)

* The derived IAT employed the nonsense syllables in place of the words that appear in parentheses

however, the formation of equivalence relations is required, then the IAT effect should only be observed for those participants who successfully complete the MTS training *and* pass the equivalence test after the IAT.

METHOD

Participants

A total of 36 single-honors psychology undergraduates took part in the experiment. None of the participants had studied stimulus equivalence or related phenomenon prior to their involvement in the study. The experiment was conducted across two 2-hr sessions on separate days of the same week. Three participants failed to return for the second day of experimentation. Two additional participants were excluded from the study because they were not fluent English speakers. A further 3 participants failed to reach criterion on the MTS training on Day 1 (see below) and their data were excluded at this point.

The 28 participants who completed the study ranged in age from 18–43 years (mode = 19 years). A total of 8 participants successfully passed the equivalence test, which was presented after the IAT. These 8 participants were categorized into the *Pass Group*. Another 8 participants who failed the equivalence test, but matched the Pass group in terms of three control conditions (i.e., MTS condition, IAT condition, and stimulus set; see below), were selected and categorized into the *Fail–Matched Group*. The data from the remaining 12 participants who also failed the equivalence test, but were not directly matched with the *Pass Group* in terms of the three conditions, were labeled the *Fail–Nonmatched Group*.

Apparatus and Stimuli

The experiment was conducted in the computer laboratory in the Department of Psychology at the National University of Ireland, Maynooth. The MTS training, IAT, and equivalence testing all were presented on Dell personal computers with Pentium 4 Processors and standard keyboards and monitors. The software used to control the presentation of stimuli and record responses was written in Microsoft Visual Basic 6. The IAT computer program was identical in all respects to the generic IAT software available for download (Greenwald, 2007). It was deemed important to use exactly the same presentation of instructions, feedback, and method of presenting stimuli (e.g., color, position on the screen, etc.) as are typically used in the now standard IAT procedure.

The words ''Spider'', ''Snake'', ''Baby,'' and ''Romance'' and three words relating to each of these but not to the other three were used to obtain a ''baseline'' IAT performance (see Table 1, Upper section). Sixteen nonsense syllables also were employed as stimuli across both the MTS training and the ''derived'' IAT procedure (see Table 1, Lower section). In addition, sixteen digital photographs were taken from the International Affective Picture System (IAPS; Lang, Bradley & Cuthbert, 1999). Four of these were pictures of spiders, four of snakes, four of babies, and four of romantic heterosexual couples. Reference numbers for the pictures used are listed in the Appendix. Two stimulus sets were constructed using the nonsense syllables and the photographs, and these are represented in Table 1. The alphanumeric code used to designate each of the nonsense syllables also is presented (participants did not see these labels). For stimulus set 1, classes 1 and 3 were assigned spider and baby functions, respectively, but for stimulus set 2 these functions were reversed. The use of these two sets rendered it unlikely that any observed IAT effect was due to some unexpected property of the nonsense syllables, rather than the predicted transfer of functions.

Participants were each asked to complete a four-item Likert-scale questionnaire that was designed to obtain their general reactions to the four stimulus categories in the IAT. Specifically, the questionnaire required participants to rate their general reaction to SPIDERS (question 1), SNAKES (question 2), BABIES (question 3) and ROMANCE (question 4). Participants responded by circling a number between 0 (labeled *Extremely Negative*) and 10 (labeled *Extremely Positive*). The number 5 along the scale was marked *No Reaction*. The questionnaire was given to participants to determine if self-reported evaluations were consistent (or perhaps inconsistent) with the observed performances on the IAT. The questionnaire was presented at the beginning and at the end of the experiment so that any changes in self-reported reactions to the four categories that might occur during the experiment (e.g., due to habituation effects) could be monitored.

Procedure

The procedure consisted of seven stages distributed across two separate two-hour sessions. The first session was held on a Tuesday and the second on Thursday of the same week. Session 1 consisted of (i) self-report reaction questionnaire, (ii) baseline IAT, (iii) MTS training; and Session 2 consisted of (iv) MTS retraining, (v) derived IAT, (vi) equivalence test, (vii) self-report reaction questionnaire. Table 2 presents a schematic overview of the stages involved in the experimental sequence.

DAY 1

The participants were each seated in front of a computer in the laboratory at the beginning of the first session. Participants were asked to refrain from talking to one another once the experiment began—participants adhered to this rule throughout the study (as monitored by two experimenters).

Self-Report Reaction Questionnaire

The self-report reaction questionnaire was distributed to each of the participants and they were asked to answer each of the questions quickly without thinking too much about the answers. The experimenters then collected the questionnaires and the participants were invited to start the IAT computer program.

Baseline IAT

The IAT program presented the following instructions across a number of display pages (the participant could move forwards or backwards through the pages using the space bar to proceed and the ''d'' key to return to the previous page):

> *INSTRUCTIONS FOR THE SORTING TASKS*
> *For each of several sorting tasks you will be shown words one at a time in the middle of the computer screen.*
> *Your task is to sort each item into its correct category as fast as you can by pressing EITHER the 'd' key or the 'k' key.*
> *IMPORTANT: Press the 'd' key using your left index finger, or the 'k' key using your right index finger.*
> *The categories associated with the 'd' and 'k' keys will be shown at the top of each screen. Please pay close attention to these category labels—they change for each sorting task!*
> *For one of the sorting tasks you will be classifying words as being either, 'snake' or 'romance'.*
> *In the other sorting task you will be classifying words as being either 'spider' or 'baby'.*
> *For each task, please judge each item on the basis of which group it appears to belong to.*
> *Please examine the next page carefully.*
> *It gives key assignment instructions for the next series of categorization trials.*
> *Press the space bar to continue.*

When the participant pressed the space bar, the display screen for the first sorting task was

Table 2

Overview of experimental sequence.

Stage	Day 1			
1	Self-report reaction questionnaire			
2	Baseline IAT *			
	Block	Function	Items assigned to left response-key	Items assigned to right response-key
	1	Practice	snake	romance
	2	Practice	spider	baby
	3	Practice	spider & snake	baby & romance
	4	Test	spider & snake	baby & romance
	5	Practice	romance	snake
	6	Practice	spider & romance	baby & snake
	7	Test	spider & romance	baby & snake
3	Matching-to-sample training **			
	Cycles of A–B, B–C, and C–D training until 24 consecutively correct responses were recorded			
	Day 2			
4	Matching-to-sample retraining			
5	Derived IAT *			
	Block	Function	Items assigned to left response-key	Items assigned to right response-key
	1	Practice	A2 (snake) ***	A4 (romance)
	2	Practice	A1 (spider)	A3 (baby)
	3	Practice	A1 & A2	A3 & A4
	4	Test	A1 & A2	A3 & A4
	5	Practice	A4	A2
	6	Practice	A1 & A4	A3 & A2
	7	Test	A1 & A4	A3 & A2
6	Equivalence Test			
	24 matching-to-sample trial-types probed D–A, C–A, and B–A relations			
7	Self-report reaction questionnaire			

* For half of the participants, the positions of Blocks 1, 3, and 4 were switched with those of 5, 6, and 7 in both Baseline and Derived IATs.

** Half of the participants received A–B, B–C, and C–D training and the other half received C–D, B–C, and A–B training.

*** Words in parentheses indicate the type of pictures that were paired with the nonsense syllables during MTS training. Note also that two different stimulus sets were constructed for the IAT (see Table 1).

presented. The specific sequence of sorting tasks, which were divided into seven blocks, differed depending on whether the participant had been assigned to the *consistent-followed-by-inconsistent* condition or the *inconsistent-followed-by-consistent* condition (hereafter referred to as consistent-first and inconsistent-first conditions, respectively). The sequence of tasks for the consistent-first condition will be described in detail.

Block 1: Snake–romance discrimination. The first sorting task presented the phrase ''Press 'd' For'' in the top-left corner and ''Press 'k' For'' in the top right corner of the computer screen. These two phrases appeared in black. Approximately 8 cm underneath these instructions the word ''snake'' appeared on the left and the word ''romance'' appeared on the right. These two words were written in green. From the participant's perspective, therefore, the instructions read ''Press d for snake'' and ''Press k for romance''. These instructions remained on the screen throughout the first block. The following additional instructions appeared before the first trial:

IF YOU MAKE AN ERROR YOU WILL SEE A RED 'X' BELOW THE STIMULUS – WHEN THIS HAPPENS, YOU HAVE TO MAKE THE CORRECT RESPONSE TO PROCEED.
THIS IS A PRACTICE TRIAL – ERRORS ARE EXPECTED.
READ THE INSTRUCTIONS, ABOVE, THEN PRESS THE SPACE BAR TO START.

When the participant pressed the space bar the additional instructions were removed immediately and 500 ms later the first stimulus was presented in the center of the computer screen. The stimulus remained on screen until the participant pressed either the 'd' or 'k' key on the computer keyboard. If a participant pressed the correct key, 'd,' given any of the snake-related target words (slither, hiss, viper) and 'k' given any of the romance-related words (dating, roses, engagement), the target was immediately removed from the screen and the next target was presented 400 ms later. If a participant pressed the incorrect or an invalid key (i.e., 'd' for a romance-related word, 'k' for a snake-related word, or any other key on the keyboard), a red 'X' immediately appeared directly underneath the target word and remained on screen. When the participant pressed the correct key both the target and the red 'X' immediately disappeared and the next target was presented 400 ms later. Each of the six target words, three snake related and three romance related, were presented randomly, without replacement, in groups of six trials for a total of 24 trials (i.e., each target was presented four times).

Immediately following the completion of trial 24, the screen cleared and performance feedback was presented to the participant. The feedback specified the percentage of correct responses and the median response time produced by the participant during the first block. The percentage of correct responses was defined as the total number of trials completed without an error divided by 24 and then multiplied by 100. The median response time was calculated across all trials, including those on which an error occurred. The response time for each trial was defined as the duration, in ms, from the presentation of the target word to the first correct response.

Immediately below the feedback message was a request for the participant to press the space bar to proceed. Upon doing so the screen cleared and the following instruction appeared: *Please examine the next page carefully. It gives key assignment instructions for the next series of categorization trials. Press the space bar to proceed.* When the participant pressed the space bar the program proceeded to Block 2 of the IAT.

Block 2: Spider–baby discrimination. Block 2 was similar to Block 1 except for the following differences. First, the two instructions at the top left and right hand corners of the screen read ''Press 'd' for Spider'' and ''Press 'k' for Baby'', respectively. Second, the words ''spider'' and ''baby'' were written in blue rather than green, and were positioned approximately 2 cm underneath the ''Press 'd''' and ''Press 'k''' phrases. Third, the additional instructions were reduced for Block 2 in that the sentence referring to errors and the red 'X' was removed (note, however, that the program treated errors for this and all other blocks in exactly the same way as in Block 1). Finally, the three spider-related words (creep, web, and tarantula) and three baby-related words (toddle, pram, milk) were presented as target stimuli.

Block 3: Consistent categories practice. This third block was similar to the previous two blocks except for the following differences. First, the instructions at the top left and top right corners of the screen were combined from Blocks 1 and 2 such that they now read ''Press 'd' for spider or snake'' and ''Press 'k' for baby or romance''. The colors of the words used in the previous blocks remained unchanged (the word ''or'' in both the left and right trials appeared in gray). Second, all 12 target words—three spider-related, three snake-related, three baby-related, and three romance-related—were presented randomly, without replacement, in two groups of 12 trials (i.e., each target was presented twice across the 24 trials).

Block 4: Consistent categories test. The fourth block was similar to Block 3, except that the first sentence of the additional instructions now read: *This is the test— Go fast, making a few errors is ok.* Furthermore, 48 trials rather than 24 were presented in a random order, without replacement, in four successive groups of 12 trials.

Block 5: Romance–snake discrimination. This block was similar to Block 1 except that the left–right positioning of the two instructions was reversed—participants now were required to press left for romance-related targets and to press right for snake-related targets. Before this block commenced, the following instructions were presented to warn the participants that the key assignments were about to change:

> *The next few blocks will change one of the categorization tasks. You will have on-screen reminders at the top throughout the block. Please use this block to remember the instruction and learn the task so you will be able to respond rapidly in the following blocks.*

Block 6: Inconsistent categories practice. Block 6 was similar to Block 3 except that the two instructions at the top left and right corners of the screen asked participants to respond to inconsistent categories; ''Press 'd' for spider or romance'' and '' Press 'k' for baby or snake''.

Block 7: Inconsistent categories test. The final block was similar to Block 4, except that the two inconsistent category instructions employed in Block 6 were used. This block also differed from others in that immediately following the last trial (i.e., trial 48), the screen cleared and the following message appeared: ''*That is the end of this part of the experiment. Please report to the experimenter''.*

Inconsistent-followed-by-consistent IAT. The procedure for the inconsistent-first condition was similar to that described above, except that the positions of Blocks 1, 3, and 4 were switched with those of 5, 6 and 7, respectively.

Matching-to-Sample Training

When a participant had completed the IAT, he or she was asked to read the following instructions (typed on a sheet of paper) while the experimenter loaded the MTS computer program:

> *Thank you for agreeing to participate in this part of the experiment, which involves responding to nonsense (unreal) words and pictures on a computer screen.*
>
> *In this part of the experiment, you will be trained to* ***match*** *nonsense words and pictures to other nonsense words. The relation between these words is not already known to you, so you will have to learn by trial and error.*
>
> *For each trial, you must look at the nonsense word and /or picture in the* ***centre*** *of the screen, which only appears briefly, and then* ***choose*** *one of the four nonsense words that appear at each corner of the screen.*
>
> *To choose one of the four words, simply press one of the four following keys on the keyboard that corresponds to your choice: the '****R****' key for the top-left word; '****U****' for top-right; '****C****' for bottom-left; and '****N****' for bottom-right. The computer will tell you whether you have made the correct choice or not. Remember, your task is to match the centre word and/or picture with one of the four words appearing just after it.*
>
> *All further instructions will be presented by the computer.*
>
> *Thank you again, and GOOD LUCK!*

On each trial during the MTS training, a sample stimulus appeared in the center of the computer screen. After 1500 ms the sample was removed and 50 ms thereafter four comparison stimuli appeared, one in each corner of the computer screen. The comparison stimuli remained visible until the participant pressed one of the four designated response keys. If the participant chose a comparison stimulus that was deemed correct by the computer program, the comparison stimuli were removed and the word ''Correct'' appeared in the middle of the computer screen where it remained for 1000 ms. A soft computer-generated chime also was presented with this visual feedback. If the participant chose a comparison stimulus that was deemed incorrect by the program, a similar sequence occurred except that the word ''Wrong'' was presented and no chime occurred. When the feedback message (''Correct'' or ''Wrong'') was removed, the screen remained blank for a 2000-ms intertrial interval. Immediately thereafter the next trial was presented.

All participants were trained using 12 MTS trial-types designed to establish four 4-member equivalence classes. Four of the trial-types involved using a complex sample stimulus composed of nonsense words and pictures; these will be described in the next subsection. The training required between 20 and 100 min (approximately) depending on each participant's performance. The training protocol was of a linear design and a schematic representation of the 12 MTS trial-types is presented in the upper section of Table 3. If, for example, the nonsense word A1 was presented as a sample, pressing the key that corresponded to the comparison B1 produced the ''Correct'' feedback, but pressing the keys that corresponded to B2, B3, or B4 produced the ''Wrong'' feedback. There were two MTS conditions—half of the participants were trained on the four A–B trial-types followed by the four B–C trial-types and then the four C–D trial-types, whereas others were trained in the reverse order; C–D, followed by B–C, followed by A–B (i.e., the order of A–B, B–C, and C–D training was counterbalanced across participants).

The following procedure was employed with those participants who were exposed to A–B followed by B–C, and C–D training. The four A–B trials were presented randomly, without replacement, in groups of four trials, such that each trial-type was presented once within each group of four trials. The position of the four comparison stimuli was randomized across trials, such that the correct comparison could appear with equal probability in any of the four corners of the computer screen. Each participant was required to produce eight consecutively correct trials in order to complete the A–B training. When A–B trials had been successfully completed, participants were exposed to B–C training and then C–D training. In both cases, the training was conducted in the same manner as the A–B training except that for the B–C training the sample stimuli consisted of B1, B2, B3, and B4, and the comparison stimuli consisted of C1, C2, C3, and C4; and for the C–D training the sample stimuli consisted of C1, C2, C3, and C4, and the comparison stimuli consisted of D1, D2, D3, and D4.

Participants were exposed to cycles of A–B, followed by B–C and C–D training until they produced 24 consecutively correct responses across eight A–B, eight B–C, and eight C–D trials. That is, each participant was required to complete the A–B, B–C, and C–D training without error. If a single error occurred during the training of any of the three conditional discriminations, the participant was returned to A–B training followed by B–C and C–D training, and so on, until 24 consecutively correct trials were recorded. When the training criterion was met the computer screen cleared and a message appeared that invited the participant to take a short break before pressing the space bar to continue. When a participant pressed the space bar, he or she was reexposed to the same repeating cycle of A–B, B–C, and C–D MTS training until another 24 consecutively correct responses were emitted. Multiple training exposures were employed to ensure that the conditional discriminations were firmly established in the behavioral repertoires of the participants. When a participant successfully completed the second cycle of training, the screen cleared and a message appeared informing the participant that this part of the experiment was complete and to report to the experimenter. This marked the end of experimentation for Day 1.

The MTS training procedure for the participants who were exposed to C–D training followed by B–C and A–B training was similar to that described above, except that the cycle of training commenced with C–D trial types and then progressed to B–C, and then to A–B trial types. In this case, therefore, if a participant failed to produce 24 consecutively correct trials at the end of the A–B training, he or she was returned to C–D training followed by B–C and A–B training, and so on, until 24 consecutively correct trials were recorded.

The complex A stimuli. As indicated previously, four of the trial types involved using a complex sample stimulus composed of nonsense words and pictures. The relevant stimuli and pictures are represented in Table 1. The four A stimuli, A1, A2, A3, and A4 consisted of four nonsense syllables, each of which was also compounded, across training trials, with four different pictures of spiders, snakes, babies, and romance, respectively. The across-trial stimulus compounding was achieved using the following procedure. On the first A–B training trial the sample stimulus consisted of a nonsense syllable located in the middle of the upper edge of an appropriate picture (e.g., JEP with a picture of a spider). On the second A–B training trial the sample consisted of a nonsense word alone (e.g., POF without the accompanying picture of a snake). On the third A–B training trial the sample consisted of a picture alone (e.g., a picture of a baby without the TOB nonsense syllable). Every subsequent cycle of three A–B training trials repeated this pattern (picture–syllable; syllable alone; and picture alone) for all exposures to the A–B training. In effect, trials with syllable alone, picture alone, and syllable–picture compounds were presented equally often across trials. For each A–B trial on which a picture was presented, with or without a nonsense syllable, the computer selected randomly from a pool of four relevant pictures (e.g., from four pictures of a snake).

The across-trial compounding procedure served to pair each of the A-stimulus nonsense words with one of the four categories of pictures (spiders, snakes, babies, and romance), and it also ensured that participants learned to match the reinforced B stimuli to the two components of the appropriate complex A stimuli (e.g., participants learned to match VIR to JEP and/or pictures of spiders). It is important to note that although the picture stimuli were paired with the A nonsense syllables via compounding, and the B nonsense syllables via matching, neither the C nor the D nonsense syllables were directly compounded or matched to the pictures during the MTS training. Thus, any transfer of evaluative functions that might occur from the pictures to the C and D stimuli, and thereby produce an IAT effect, could not be explained in terms of direct associative processes.

Day 2

When participants were again seated in front of their computers they were reminded to refrain from talking to each other once the experiment began. They were then informed that they were to be provided with 'refresher' training on the matching task that they had completed during the previous session.

Matching-to-Sample Retraining

The MTS retraining was exactly the same as the MTS training during the first session (two separate blocks of 24 trials correct were required). The retraining was included to ensure that the matching performances were firmly established immediately prior to a participant's exposure to the next phase.

Derived IAT

The derived IAT was the same as the baseline IAT in all respects except that the nonsense syllables employed in the MTS training and retraining were substituted for the real words. Table 1 provides a complete breakdown of the exact pattern of substitutions among the real and nonsense words between the baseline and derived IATs. As can be seen from the Table, each of the four A stimuli were employed as category labels (for set 1, JEP, POF, TOB, and MIQ replaced *spider, snake, baby,* and *romance,* respectively), and the B, C, and D stimuli were used as target words (e.g., for set 2, VIR, PUK, and ROG replaced *toddle, pram,* and *milk,* respectively). Thus, for example, a consistent-categories IAT practice or test trial, using Set 1 stimuli, read ''Press 'd' for JEP or POF'' and ''Press 'k' for TOB and MIQ'' and the 12 remaining nonsense syllables (e.g., VIR, BOC, etc.) were presented as target words. The derived IAT thus presented participants with a set of categorization tasks that were functionally similar to the baseline IAT, assuming that all of the nonsense stimuli had acquired the relevant evaluative functions (e.g., Spider for A1, B1, C1, and D1; Snake for A2, B2, C2, and D2; and so on).

Equivalence Test

Immediately following the derived IAT, participants were asked to read the following instructions (typed on a sheet of paper) while the experimenter loaded the equivalence testing program:

> *In this part of the experiment, you must look at the nonsense word in the* **centre** *of the screen, which only appears briefly, and then* **choose** *one of the four nonsense words or four pictures that appear at each corner of the screen.*
> *To choose one of the four words, simply press one of the four following keys on the keyboard that corresponds to your choice: the '**R**' key for the top-left word; '**U**' key for top-right; '**C**' key for bottom-left; and '**N**' key for bottom-right.*
> *During this stage the computer will not tell you whether you have made the correct choice or not. However, you can still get all the tasks correct. Do your best to get everything right.*
> *All further instructions will be presented by the computer.*
> *Thank you again, and GOOD LUCK!*

The MTS procedure for the equivalence test was similar to that used during the MTS training and retraining with two key differences. First, the test consisted of 24 MTS trial types, which probed for the four D–A, four C–A, and four B–A relations. Although these test probes constitute only a subset of those that could have been included, they do provide combined tests for equivalence relations (Sidman, 1994). Furthermore, given the already protracted procedure, there was a danger that a lengthy equivalence test would increase the possibility of obtaining false negative results due to general fatigue and/or distraction effects. Probing only a key subset of emergent relations reduced this possibility.

Of the 24 MTS test trials, 12 presented the nonsense syllables alone, and 12 presented the pictures alone, from the complex A stimuli. The pictures-alone trials functioned as a test for the transfer of valence functions, because it required that participants match the appropriate pictures to the B, C, and D stimuli. Appropriate picture matching served to indicate that the relevant valences had transferred to the related stimuli, and thus the equivalence classes were not simply composed of nonsense syllables with no valence functions. As was the case during the training, the computer selected randomly from the pool of four pictures available to it for each A stimulus. The full list of trial types is presented in the lower section of Table 3.

Table 3

A schematic representation of the trained conditional discriminations and tested equivalence relations.

Sample	Correct comparison	Incorrect comparison
Trained conditional discriminations *(The A stimuli consisted of picture-syllable compounds, syllables alone, and pictures alone, presented across a three-trial repeating cycle)*		
A1	B1	B2, B3, B4
B1	C1	C2, C3, C4
C1	D1	D2, D3, D4
A2	B2	B1, B3, B4
B2	C2	C1, C3, C4
C2	D2	D1, C3, D4
A3	B3	B1, B2, B4
B3	C3	C1, C2, C4
C3	D3	D1, D2, D4
A4	B4	B1, B2, B3
B4	C4	C1, C2, C3
C4	D4	D1, D2, D3
Tested equivalence relations *(12 trials presented syllables alone and 12 presented pictures alone as A stimuli)*		
D1	A1	A2, A3, A4
D2	A2	A1, A3, A4
D3	A3	A1, A2, A4
D4	A4	A1, A2, A3
C1	A1	A2, A3, A4
C2	A2	A1, A3, A4
C3	A3	A1, A2, A4
C4	A4	A1, A2, A3
B1	A1	A2, A3, A4
B2	A2	A1, A3, A4
B3	A3	A1, A2, A4
B4	A4	A1, A2, A3

The second difference between the MTS training and equivalence test was that no feedback (i.e., the words ''Correct'' and ''Wrong'') was presented on any trial— instead, the program simply progressed to the intertrial interval. The MTS testing trials were presented randomly, without replacement, in a single block of 24 trials. Immediately following trial 24 the screen cleared and a message appeared inviting the participant to report to the experimenter.

Self-Report Reaction Questionnaire: End

Upon alerting the experimenter to the end of the equivalence test, each participant was asked to complete the self-report reaction questionnaire for a second time. Subsequently, the questionnaires were collected and the participants were informed that the experiment was complete and that a debriefing session would be held the following week.

RESULTS

In the current experiment 8 participants produced at least 90% correct responding on the equivalence test, which followed the derived IAT. These participants are labeled the *pass group*. Eight participants were then selected randomly from the remaining 20 who matched the pass group, on a participant-by-participant basis, in terms of the three control conditions (i.e., MTS condition, IAT condition, stimulus set); these are labeled the *fail-matched group*. The remaining participants, who also failed the equivalence test, are labeled the *fail-nonmatched group*. Information pertaining to the three groups is presented in Table 4.

Self-Report Reaction Questionnaire

The results from the self-report reaction questionnaire that was administered before and after participants completed the experiment are presented in Table 5. The questionnaire was used to determine the participants' self-rated reactions to spiders and snakes (predicted negative reactions), and to babies and romance (predicted positive reactions). To reduce the data set, the two ratings for the *spider* and *snake* questions were totaled as were the two ratings for the *babies* and *romance* questions. Visual inspection of the data indicates that all but one of the participants produced lower ratings for spiders and snakes combined than for babies and romance combined at the beginning of the experiment (participant 4 reported an equally positive reaction to both sets of stimuli). At the end of the experiment all participants produced lower ratings for spiders/snakes than for babies/romance. Due to the low n, nonparametric tests were employed to analyze the data. A series of within-participant comparisons were made using Wilcoxon Signed Ranks tests. There was a statistically significant difference in the rating between spiders/snakes and babies/romance for the pass group at the start of the experiment ($z = -2.37$, $p = .02$) and at the end of the experiment ($z = -2.52$, $p = .01$). A similar result was found for the fail-matched group at the start ($z = -2.52$, $p = .01$) and the end ($z = -2.52$, $p = .01$), and for

Table 4

MTS Condition, IAT condition, stimulus set and equivalence test performance for each participant in the pass, fail-matched, and fail-nonmatched groups.

Participant	MTS condition	IAT condition	Stimulus set	Equiv test Correct/24
		Pass group		
1	A–B first	Consistent first	1	24
2	A–B first	Consistent first	1	22
3	A–B first	Consistent first	2	24
4	A–B first	Inconsistent first	1	23
5	C–D first	Inconsistent first	2	24
6	C–D first	Inconsistent first	2	24
7	C–D first	Inconsistent first	1	24
8	C–D first	Consistent first	1	24
		Fail-matched group		
9	A–B first	Consistent first	1	11
10	A–B first	Consistent first	1	10
11	A–B first	Consistent first	2	15
12	A–B first	Inconsistent first	1	14
13	C–D first	Inconsistent first	2	12
14	C–D first	Inconsistent first	2	18
15	C–D first	Inconsistent first	1	11
16	C–D first	Consistent first	1	15
		Fail-nonmatched group		
17	A–B first	Inconsistent first	1	6
18	A–B first	Inconsistent first	1	14
19	A–B first	Consistent first	2	8
20	C–D first	Inconsistent first	1	11
21	C–D first	Inconsistent first	1	12
22	C–D first	Consistent first	1	13
23	C–D first	Consistent first	2	19
24	C–D first	Consistent first	1	3
25	A–B first	Inconsistent first	2	10
26	C–D first	Consistent first	2	14
27	C–D first	Consistent first	2	7
28	C–D first	Consistent first	1	8

the fail-nonmatched group at both the start ($z = -3.06$, $p = .002$) and end of the experiment ($z = -3.06$, $p = .002$). These results indicate that both the pass and fail groups demonstrated the predicted negative reactions to spiders/snakes relative to the predicted positive reactions to babies/romance.

The ratings data also were explored for any differences in self-report reactions that emerged among pass and fail groups. Four between-groups comparisons were made using Kruskall-Wallis tests. The results of these analyses showed no significant differences among the three groups for ratings at the start of the experiment for spiders/snakes ($H = .26$, $df = 2$, $p = .88$) or for babies/romance ($H = 1.31$, $df = 2$, $p = .52$). Likewise there was no significant difference at the end of the experiment among the three groups on ratings of either spiders/snakes ($H = .26$, $df = 2$, $p = .87$) or of babies/romance ($H = .001$, $df = 2$, $p = .99$). Thus, any differences that emerge between the groups on the subsequent measures cannot readily be attributed to different reactions to the emotive content of the stimuli.

The Implicit Association Tests

There are numerous algorithms that may be used to score the IAT, some more complex than others. For the purposes of the current research, algorithm C4, as specified by Greenwald et al. (2003), was used. Although this scoring method might not be the most sensitive measure available, Greenwald et al. suggest that it is appropriate for laboratory-based uses of the IAT with small data sets. The IAT effect is derived from response latency which is measured on each trial from the point of target onset to the first correct response emitted by the participant. Although the IAT

Table 5

Self-report ratings for each of the participants in the pass group, the fail-matched group, and the fail-nonmatched group for spiders and snakes combined, and babies and romance combined, recorded at the start and at the end of the experiment.

Pass group				
	Exp start		Exp end	
Participant	Spiders/Snake	Babies/Romance	Spiders/Snakes	Babies/Romance
1	5	17	4	19
2	14	20	16	20
3	8	16	8	15
4	7	7	6	7
5	4	14	7	13
6	6	17	5	16
7	4	14	8	14
8	10	15	11	15
Fail-matched group				
	Exp start		Exp end	
Participant	Spiders/Snake	Babies/Romance	Spiders/Snakes	Babies/Romance
9	8	12	7	14
10	8	18	8	18
11	3	12	5	12
12	4	18	4	17
13	9	16	8	16
14	9	15	9	14
15	9	19	10	19
16	8	20	8	14
Fail-nonmatched group				
	Exp start		Exp end	
Participant	Spiders/Snake	Babies/Romance	Spiders/Snakes	Babies/Romance
20	11	16	12	18
21	12	14	10	15
22	8	10	6	14
23	8	16	10	17
24	5	17	5	18
25	5	12	5	11
26	6	11	3	15
27	9	13	8	13
28	9	15	9	15

effect is derived from response latency, it also reflects accuracy because latencies are typically longer on error trials because the first correct response always follows at least one incorrect response.

The C4 algorithm involves first recoding each response latency for each participant on every trial. To control for spurious outliers due to inattention or occasional distraction, all latencies shorter than 300 ms are recoded to 300 ms, and all latencies longer than 3000 ms are recoded to 3000 ms. A mean latency for each participant is then calculated from the recoded data for the consistent categories practice block and the consistent categories test block. The two means are then summed and divided by 2 to produce a single mean latency for both consistent category blocks for each participant. The same approach is taken with the inconsistent categories test and practice blocks, which again produces a single mean latency for the inconsistent blocks. A longer mean latency for the inconsistent versus consistent blocks suggests an IAT effect. (It is standard practice in IAT research to counterbalance the consistent- and inconsistent-first conditions across participants. The effects of such counterbalancing have been investigated systematically by IAT researchers, and although IAT effects tend to be slightly larger when a consistent condition precedes an inconsistent condition these differences tend not to be statistically significant [Greenwald et al., 1998].)

In the current experiment it was expected that all participants would produce an IAT effect in the baseline IAT because it used natural language categories, and all participants produced the expected positive and negative self-report reactions to these categories averaged across the two self-reports. The prediction was different for the derived IAT, however. If derived equivalence relations function like semantic relations, then the pass group should produce reliable IAT effects with the derived IAT. For the fail groups, however, the difference between consistent and inconsistent categorization tasks should vary randomly, and thus the number of participants showing an IAT effect within the group should not rise significantly above chance. To support the prediction for the pass group, all 8 participants would have to produce the predicted IAT effect. That is, we assumed that each participant has a 0.5 probability of producing an IAT effect by chance (excluding the highly unlikely result that the overall mean latencies are exactly the same for both consistent and inconsistent blocks). Based on this assumption, 4 of the 8 pass-group participants would be expected to produce an IAT effect, but the probability of all 8 participants producing this effect would be .03 (i.e., 0.5^4), which is statistically significant at the .05 level. In applying the same statistical assumptions to the two fail groups, all 8 participants in the matched group, and 10 out of the 12 in the nonmatched group, would have to produce an IAT effect to be statistically significant ($p < 0.05$).

In examining the data obtained in the current study, the key predictions were upheld (see Table 6). All of the participants in the pass, fail–matched, and fail–nonmatched groups produced shorter mean latencies for the consistent categorization tasks than for the inconsistent tasks in the baseline IAT. In the derived IAT a similar pattern was observed for the pass group—again, all 8 participants produced a clear IAT effect ($p = .03$). In the fail-matched group, however, only 2 of 8 participants produced shorter latencies for consistent versus inconsistent tasks ($p = .125$) and in the fail–nonmatched group 8 of the 12 participants produced shorter latencies for consistent versus inconsistent tasks ($p = .125$).

Additional statistical analyses were then conducted on the actual response latency data (for ease of comparison, the mean response latencies presented in Table 6 have been averaged and are presented graphically in Figure 2). First, for the baseline IAT, a series of Wilcoxon Signed Ranks tests showed a statistically significant difference between consistent and inconsistent tasks for the pass group ($z = -2.52$, $p = .01$); for the fail–matched group ($z = -2.52$, $p = .01$); and for the fail–nonmatched group ($z = -3.06$, $p = .002$). In the derived IAT, however, the difference between consistent and inconsistent tasks was statistically significant for the pass group ($z = -2.52$, $p = .01$) but for both the fail–matched group ($z = -1.26$, $p = .21$) and the fail–nonmatched group ($z = -1.49$, $p = .14$) the difference did not reach statistical significance.

As predicted, therefore, the pass and fail groups produced the typical IAT effect in the baseline IAT, but only the pass group showed this effect with the derived IAT. These results support the prediction that derived equivalence relations function like semantic categories in natural language. In effect, the data suggest that for the pass group, but not for the fail groups, all four nonsense stimuli in each of the four classes acquired spider, snake, baby, and romance functions, respectively.

Directly and Indirectly Related Stimuli in the Derived IAT

The data described above would appear to support the prediction that evaluative functions of spiders, snakes, babies, and romance had transferred, for the pass group, from the pictorial stimuli to each of their related class members. Additional analyses are required, however, to determine if this was in fact the case. The MTS training ensured that the B stimuli were directly paired with the pictorial A stimuli, whereas the C and D stimuli were only indirectly related (via B) to A. To determine if the direct versus indirect matching of stimuli had a differential impact in producing the derived IAT effect, the latencies for B, C, and D stimuli were isolated and analyzed separately. Table 7 presents the mean response latencies for the B, C, and D stimuli for the pass,

Table 6

Adjusted mean response latencies for each of the participants in the pass group, fail–matched group, and fail–nonmatched group for the baseline and derived IAT for the consistent and inconsistent categorization tasks.

Pass group				
	Baseline IAT		Derived IAT	
Participant	Consistent	Inconsistent	Consistent	Inconsistent
1	773	1352	965	1353
2	682	1172	1334	1639
3	610	1023	828	1548
4	530	906	734	1602
5	539	803	624	1456
6	704	1138	860	1451
7	638	1002	652	1125
8	776	1073	968	1279
Fail–matched group				
	Baseline IAT		Derived IAT	
Participant	Consistent	Inconsistent	Consistent	Inconsistent
9	572	775	914	959
10	668	840	2682	2665
11	574	983	2351	1577
12	955	1273	1285	843
13	643	1103	1463	2234
14	764	980	1294	1000
15	662	954	1265	1036
16	788	955	1858	1624
Fail–nonmatched group				
	Baseline IAT		Derived IAT	
Participant	Consistent	Inconsistent	Consistent	Inconsistent
17	627	1046	1024	1109
18	541	783	1015	1095
19	983	1323	1332	1751
20	1115	2697	1690	2193
21	887	981	1762	1747
22	674	1086	1499	1202
23	875	1032	1442	1335
24	744	1158	1658	1605
25	907	1057	1510	1523
26	829	1462	1289	1655
27	713	1085	1294	1359
28	799	1181	860	1167

fail–matched, and fail–nonmatched groups with the derived IAT. It should be noted that some IAT researchers have recommended against analyzing the IAT effect for individual target stimuli (De Houwer, 2001). Nevertheless, it was deemed appropriate in the current context because we were not seeking to assess the differential associative strengths of each target within a category, but simply to determine if the IAT effect was obtained for each target for the pass but not for the fail groups.

Table 7 shows that all 8 participants in the pass group produced shorter response latencies for the consistent categorization tasks than for the inconsistent tasks for each of the B, C, and D stimuli (if we apply the same statistical logic to the individual stimuli that were applied to the overall IAT effect, p = .03 for each stimulus). In contrast, neither the fail–matched nor the fail–nonmatched groups demonstrated this reliable pattern (p > .05 for each stimulus for each group).

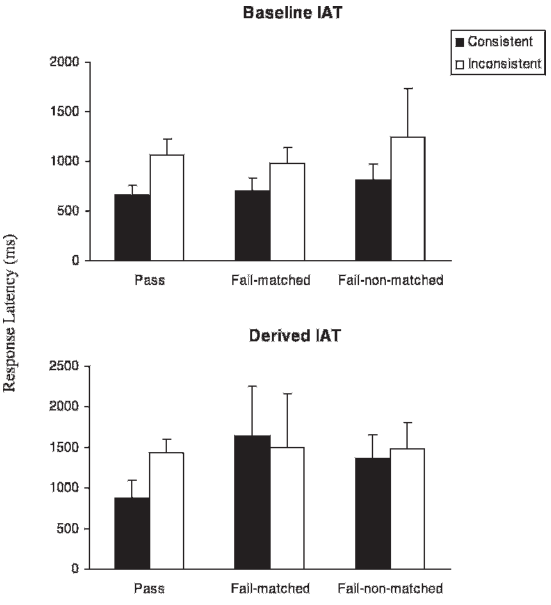

Fig. 2. Overall mean response latencies for the pass group, the fail–matched group, and the fail–nonmatched group for consistent and inconsistent tasks on both the baseline (top panel) and derived (bottom panel) IAT.

For ease of comparison, the average latencies for the three groups are presented graphically in Figure 3. The data for the B, C, and D stimuli were analyzed using a series of Wilcoxon Signed Rank tests. For the pass group, results showed a statistically significant difference between consistent and inconsistent tasks for the B stimuli ($z = -2.52$, $p = .01$), the C stimuli ($z = -2.52$, $p = .01$), and the D stimuli ($z = -2.52$, $p = .01$). For the fail–matched group, however, the difference between consistent and inconsistent tasks did not reach statistical significance for any of the three stimuli; B ($z = -1.12$, $p = .26$), C ($z = -.98$, $p = .33$) and D ($z = -1.12$, $p = .26$). This pattern was also observed for the fail–nonmatched group; B ($z = -1.49$, $p = .14$), C ($z = -1.49$, $p = .14$) and D ($z = -1.02$, $p = .31$). These results indicate that for the pass group, both directly related (B) and indirectly related (C and D) stimuli produced the typical IAT effect. It appears, therefore, that for the pass group, the positive and negative evaluative functions of the pictorial A stimuli transferred to all equivalence class members whether directly or indirectly related. For the fail groups, however, neither directly nor indirectly related stimuli produced the typical IAT effect and thus there was no evidence of a derived transfer of evaluative functions.

DISCUSSION

The current study demonstrated that the IAT may be used to assess a derived transfer of functions in accordance with four equivalence relations. The results are particularly noteworthy because there is only one preliminary piece of research in the literature that assessed derived relations with the IAT (Barnes-Holmes et al., 2004), and this research did not employ the generic IAT as developed by Greenwald and his associates (Greenwald, 2007; Greenwald et al., 1998; Greenwald et al., 2003)—e.g., there was no self-correction response required. Furthermore, the previous research employed equivalence relations composed purely of nonsense syllables with no emotive stimuli, and thus the reported IAT effect was not based on the juxtaposition of emotionally disparate categories. Instead, the effect emerged from the requirement to categorize different equivalence classes together. The previous research therefore failed to assess a derived transfer of emotive functions, which is perhaps the most important feature of the IAT—its sensitivity to emotionally valenced stimuli. The current study, however, presents a more complete test of the IAT as a measure of derived transfer.

The present research also is important because it clearly demonstrates that the IAT effect can be obtained before participants are exposed to a formal MTS equivalence test. In fact, the data show that only those participants who subsequently passed the equivalence test produced a reliable IAT effect. In the previous study reported by Barnes-Holmes et al. (2004), all of the participants were trained and tested for equivalence class formation before being exposed to the IAT task, and thus it was not possible to determine if the MTS training or testing, or perhaps both, were necessary to produce the IAT effect. The current findings indicate that successfully completing the MTS training (even four times across two separate sessions) is not sufficient to produce the predicted outcome on the IAT—participants must also demonstrate the formation of derived equivalence relations to generate the IAT effect (see Barnes-Holmes et al., 2005, for a similar finding in the context of semantic priming).

Table 7

Mean response latencies for each of the participants in the pass group, the fail–matched group, and the fail–nonmatched group for the B, C, and D stimuli for consistent and inconsistent categorization tasks.

Pass group						
	B Stimuli		C stimuli		D Stimuli	
Participant	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent
1	914	1284	1177	1423	803	1353
2	1293	1573	1244	1718	1466	1627
3	855	1561	818	1442	812	1644
4	737	1462	724	1735	744	1609
5	647	1606	650	1591	576	1171
6	881	1377	890	1421	808	1554
7	639	1226	662	1010	657	1141
8	922	1179	951	1397	1030	1263
Fail–matched group						
	B Stimuli		C stimuli		D Stimuli	
Participant	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent
9	881	982	849	962	1011	934
10	2588	2463	2649	2747	2808	2785
11	2060	1324	2605	1798	2388	1607
12	1103	798	1383	886	1369	844
13	1001	1845	1363	2298	2024	2559
14	1030	1015	1560	1034	1293	953
15	1196	988	1429	942	1171	1170
16	1858	1270	1895	1754	1821	1846
Fail–nonmatched group						
	B Stimuli		C Stimuli		D Stimuli	
Participant	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent
17	933	1046	1106	1063	1032	1217
18	961	1094	1032	1231	1053	961
19	1275	1785	1306	1822	1415	1645
20	1805	2418	1628	1996	1637	2165
21	1709	1755	1746	1880	1831	1607
22	1661	1234	1524	1117	1312	1254
23	1487	1295	1245	1122	1595	1587
24	1722	1644	1699	1709	1553	1461
25	1632	1670	1284	1533	1614	1366
26	1432	1644	1179	1680	1255	1642
27	1154	1298	1348	1211	1380	1569
28	986	1215	800	1132	794	1155

The current data also show that direct stimulus associations are not necessary to produce the IAT effect. All of the participants in the current study were exposed to the IAT before an equivalence test, and thus the C and D stimuli had not been directly paired with the emotionally valenced pictures (or the compounded nonsense syllables). Statistical analyses indicated, however, that the IAT tasks involving the C and D stimuli produced significant differences across consistent versus inconsistent tasks. These differences provide evidence to support the argument that the IAT effect can be produced via a transfer of valence functions through derived relations before a formal MTS test, and this result is consistent with previous research in the behavioral literature (e.g., Barnes & Keenan, 1993; Barnes-Holmes, Keane, Barnes-Holmes, & Smeets, 2000; Dougher et al., 1994; Roche & Barnes, 1997).

Of the 28 participants who completed the conditional discrimination training, only 8

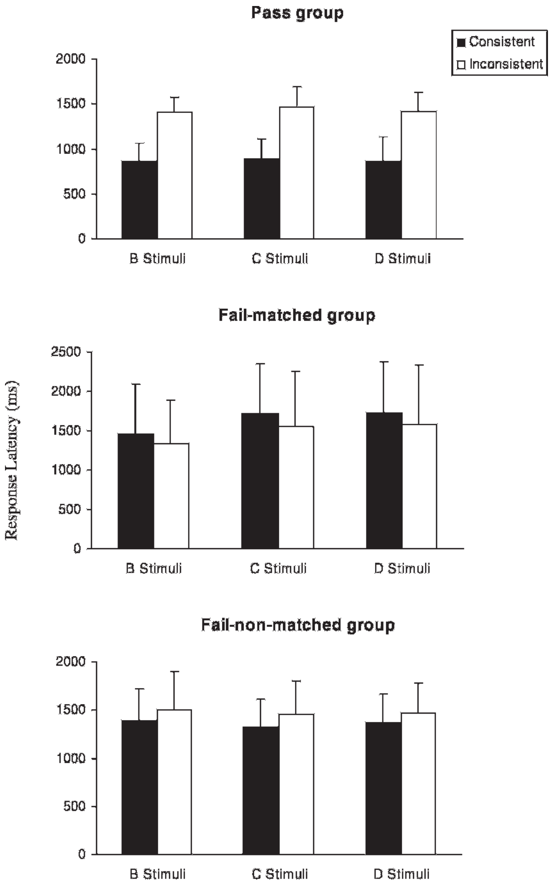

Fig. 3. Overall mean response latencies for the pass group (upper panel), the fail–matched group (middle panel), and the fail–nonmatched group (lower panel) for the B, C, and D stimuli for both consistent and inconsistent tasks.

passed the equivalence test (following the IAT), which could be seen as a relatively low yield. On balance, the number and size of the equivalence relations, and the training and testing protocol that was employed, do not typically produce high numbers of passes on equivalence tests (e.g., Barnes-Holmes et al., 2005). For example, the trained and tested stimulus relations involved four 4-member equivalence classes (or perhaps five members if the picture–nonsense syllable compounds are treated as separate stimuli), and participants were provided with only a single exposure to a test that probed for equivalence relations alone (i.e., without individual probes for symmetry and transitivity). Furthermore, the experimental protocol was simultaneous, in that all training trials were presented before the test trials (see Fields et al., 1997). Consequently, the number of participants who passed the equivalence test in the current study was consistent with previous research and indeed was more or less expected. In fact, we deliberately employed an experimental design that would produce a range of ''passes'' and ''fails'' in order to compare these groups' performances on the IAT.

A related issue concerns the fact that all participants were presented with only one exposure to each trial of the equivalence test following the IAT. Given the correlational nature of the current study it was deemed essential that the predicted equivalence relations in the pass group should be observed immediately after the IAT. Typically, in studies of derived stimulus relations participants are exposed to multiple cycles of training and/or testing, and this practice usually increases the number of successful test performances (e.g., Barnes & Keenan, 1993; Devany, Hayes, & Nelson, 1986). In the current study, however, exposing participants to extended test exposures following the IAT was deemed redundant in terms of addressing the core research question. Specifically, if an IAT effect had been produced via the formation of a set of derived equivalence relations, such relations should, in principle, produce reliable equivalence test performances immediately after the IAT (and all 8 members of the ''pass group'' produced exactly this pattern). If, however, equivalence test performances emerged only gradually after the IAT, such data would not provide strong support for the argument that the previously observed IAT effect was produced by those same derived relations. In this sense, therefore, the brief equivalence test employed in the current study constitutes a very strong test of the core hypothesis.

The fact that the IAT effect was generated with stimuli that were not directly associated with the emotionally valenced pictures raises some interesting questions concerning the IAT effect itself. All of the participants successfully completed the MTS training (four times) and thus each of the stimuli had been repeatedly paired in a linear chain (Picture/A–B–C–D). If the formation of a linear associative chain such as this was sufficient to produce the IAT effect, then all of the participants should have shown the effect—but they did not. Instead, the subsequent demonstration of derived equivalence relations seemed to be required, and thus the IAT appears to measure not just simple associations, but derived or verbal relations. Critically, within the behavioral tradition, derived relations have attracted increasing attention because they appear to provide a behavior-analytic model of semantic or verbal control in natural language (e.g., Barnes-Holmes et al., 2005; see Hayes, Barnes-Holmes, & Roche, 2001; Sidman, 1994, for reviews).

The current finding that an IAT effect only emerged for those participants who also demonstrated derived or verbal relations could be seen as broadly consistent with very recent findings within the IAT literature. In two separate studies, IAT researchers have shown that it is possible to generate an IAT effect simply by instructing participants about the positive and negative valence of nonsense stimuli subsequently employed in an IAT (De Houwer, 2006; Gregg, Seibt, & Banaji, 2006). As a result, De Houwer concluded that direct stimulus pairings or associations are not necessary to produce an IAT effect, and that ''IAT effects can be based on conscious propositional knowledge.'' (p. 184). If derived stimulus relations are assumed to provide a behavior-analytic model of the basic units of ''conscious propositional knowledge,'' the current data appear to support De Houwer's conclusion. As an aside, it should be noted that the previous studies by De Houwer and Gregg et al. involved instructing participants that nonsense syllables with specific formal features possessed particular functions (e.g., all targets containing *Niff* have positive characteristics and all targets containing *Lup* have negative characteristics). Consequently, the IAT effects observed could have been produced in part through primary stimulus generalization rather than through the formation of arbitrary stimulus relations, as was the case in the present study.

In attempting to explain the derived IAT effects observed in the current research it might be tempting to appeal to some form of higher-order classical conditioning. In mounting such an argument, however, it is important to note that a delayed MTS procedure was employed in the current study and the sample and comparison stimuli were not present simultaneously on the computer screen. Consequently, the trained relations were unidirectional, and thus the transfer effects were predicted based on derived stimulus relations rather than forward Pavlovian conditioning per se. In effect, the MTS training could be conceptualized as a linear chain of stimulus pairings in which the emotionally valenced A stimuli predicted B, which predicted C, which predicted D. Critically, a transfer of functions from the A to the C and D stimuli cannot occur through forward conditioning because the C and D stimuli followed, rather than preceded the presentation of the B and A stimuli—forward Pavlovian conditioning requires that the conditioned stimulus predicts the subsequent presentation of the unconditioned stimulus.

On balance, one might appeal to a combination of backward and higher-order or sensory preconditioning to explain the current data. Although associative learning theorists have reported so-called backward sensory preconditioning (Ward-Robinson & Hall, 1996), these effects have been explained in terms of mediated forward conditioning (Hall, 1996; Urcuioli, 1996). According to this argument, training an A–B relation may cause A to generate an internal representation of B, and then if an A–C relation is trained, a forward associative chain is established in which A activates the internal representation of B, which is then paired with C. Given this mediated associative chain, acquiring a new B–C relation should be facilitated, and this is what researchers have observed (Nakagawa, 2005; see Hall, 1996; Urcuioli, 1996, for reviews).

The linear design of the current MTS training, however, does not permit the type of mediated forward conditioning outlined above. Indeed, Hall (1996) argued that the current design should fail to produce the type of transfer effects observed in the current study:

> Associative links formed in the first stage of training… might allow A1 and A2, when presented as comparison stimuli in the test, to evoke representations of B1 and B2. But the new sample stimulus (C1) would be able to evoke the representation of the trained sample (A1 for the choice between B1 and B2) only by way of a chain of backward associations (i.e., C–B–A). Such backward associations are not readily formed… In different terminology, these training procedures do not establish

> the symmetry relation, and, hence, the equivalence test—which depends on the effectiveness of this relation—will be failed (p. 248).

Parenthetically, Hall recognized the evidence for backward associative conditioning but also pointed out that such conditioning occurs ''only in a rather restricted set of conditions… and these conditions are not especially well met'' (p. 238) in the transfer studies that he reviewed.

Of course, alternative classical conditioning accounts might still be offered for the current data, but it seems important to acknowledge a possibly important role for the verbal histories of the current participants. Indeed, some behavior analysts (e.g., Hayes et al., 2001; Leader, Barnes, & Smeets, 1996) and associative learning theorists (e.g., Lovibond & Shanks, 2002) have argued that human verbal abilities often serve to ''mediate'' the effects of respondent contingencies on verbally able humans. In fact, these authors have argued that verbal responding may be necessary in order to observe the types of human respondent conditioning effects that have been reported in the literature (see also Lovibond, 2003). Overall, therefore, it would be unwise to assume, without question, that the derived IAT effects observed in the current study were simply due to traditional classical conditioning processes.

The present findings indicate that the IAT may be used successfully to assess a derived transfer of functions, but many questions remain. For example, the emotive stimuli employed in the current study consisted of pictures. Although these pictures were standardized IAPS materials, and participants rated their emotional valence in accordance with experimental predictions, it is still possible that the IAT effect emerged in part from the non-emotive categorization of the stimuli (e.g., snakes and spiders = nonhuman : babies and romance = human). Nevertheless, even if non-emotional functions were involved, the current data still demonstrate that the IAT may be a useful method for assessing a derived transfer of functions (emotive or otherwise).

Another question arising from the current work pertains to the IAT itself. The IAT was not designed by behavior-analytic researchers and thus very little is known about the behavioral principles underlying its effect. Addressing this issue will be both complex and difficult because IAT studies often differ in minor but perhaps important procedural details (e.g., number of trials per block, type of stimuli, level of performance feedback), and a variety of scoring algorithms have been used to transform and analyze IAT data (Greenwald et al., 2003). A complete behavior-analytic investigation of the IAT is thus beyond the scope of any single study. Nevertheless, the current research has demonstrated that the IAT is sensitive to a derived transfer of functions using a standard procedure and a standard scoring algorithm. As such, the current work constitutes a first important step in providing a behavior-analytic understanding of the ubiquitous IAT effect.

REFERENCES

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and non-arbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior, 59*, 61–81.

Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. M. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record, 50*, 493–512.

Barnes-Holmes, D., Staunton, C., Barnes-Holmes, Y., Whelan, R., Stewart, I., Commins, S., et al. (2004). Interfacing relational frame theory with cognitive neuroscience: Semantic priming, the implicit association test, and event related potentials. *International Journal of Psychology and Psychological Therapy, 4*, 215–240.

Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., Walsh, D., et al. (2005). Derived stimulus relations, semantic priming and event-related potential: Testing a behavioral theory of semantic networks. *Journal of the Experimental Analysis of Behavior, 84*, 417–434.

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. (2003). A structural analysis of indirect measures of attitudes. In J. Musch, & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J. (2006). Using the implicit association test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Leaning and Motivation, 37*, 176–187.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243–257.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*, 331–351.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., et al. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67*, 367–389.

Greenwald, A. G. (2007). *Implicit association test materials.* Retrieved April 10, 2007, from University of Washington web site: http//faculty.washington.edu/agg/iat_materials.htm

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*, 1–20.

Hall, G. (1996). Learning about associatively activated stimulus representations: Implications for applied equivalence and perceptual learning. *Animal Learning & Behavior, 24*, 233–255.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

Hayes, S. C., Kohlenberg, B. S., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior, 56*, 119–137.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Instruction manual and affective ratings.* (Tech. Rep. No. A-4). Gainesville, FL: University of Florida, Centre for Research in Psychophysiology.

Leader, G., Barnes, D., & Smeets, P. M. (1996). Establishing equivalence relations using a respondent–type training procedure. *The Psychological Record, 46*, 685–706.

Lovibond, P. F. (2003). Casual beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 97–106.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology, 28*, 3–26.

Mitchell, C. J., Anderson, N. E., & Lovibond, P. F. (2003). Measuring evaluative conditioning using the Implicit Association Test. *Learning and Motivation, 34*, 203–217.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.

Nakagawa, E. (2005). Emergent, untrained stimulus relations in many-to-one matching-to-sample discrimination in rats. *Journal of the Experimental Analysis of Behavior, 83*, 185–195.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus functions in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67*, 275–301.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative Inc.

Smyth, S., Barnes-Holmes, D., & Forsyth, J. (2006). A derived transfer of simple discrimination and self-report arousal functions in spider fearful and non-spider fearful participants. *Journal of the Experimental Analysis of Behavior, 85*, 223–246.

Urcuioli, P. J. (1996). Acquired equivalences and mediated generalization in pigeon's matching-to-sample. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 55–70). Holland: Elsevier Science.

Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 22*, 395–404.

Wulfert, E., & Hayes, S. C. (1988). Transfer of conditional order response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125–144.

Received: October 31, 2006
Final acceptance: May 3, 2007

APPENDIX

Reference numbers for the International Affective Pictures System (IAPS) pictures used in the experiment

Spider Images	Snake Images
1200	1090
1201	1052
1205	1114
1220	1050
Baby Images	Images of Romantic Couples
2070	2550
2040	4599
2071	4601
2080	4610