# COMPETING CONTINGENCIES OVER DERIVED RELATIONAL RESPONDING: A BEHAVIORAL MODEL OF THE IMPLICIT ASSOCIATION TEST

Amanda Gavin

*University of Teesside, UK*

Bryan Roche

*National University of Ireland, Maynooth*

Maria R. Ruiz

*Rollins College, Florida*

*Subjects were exposed to a word-picture association training phase in which each of 2 arbitrary nonsense syllables printed in blue and red font, respectively, were paired with either sexual or aversive photographic images. Subjects were then exposed to an equivalence training procedure that led to the formation of 2 3-member equivalence relations, each containing 1 of the 2 nonsense syllables in their respective color fonts, and 2 novel nonsense syllables in black font. In effect, equivalence class 1 (blue) was associated with sexual images, while equivalence class 2 (red) was associated with aversive images. Subjects were then exposed to a 2-block test in which sexual and aversive images and all members of the trained equivalence relations, presented in black font, were employed. In 1 block, subjects were instructed to produce responses that were compatible with their laboratory history. Specifically, subjects were instructed to produce the same operant response on a computer keyboard upon the presentation of both sexual images and members of equivalence class 1 (blue), and to produce another common response upon the presentation of aversive images and members of equivalence class 2 (red). In the second block of the test the instructions were juxtaposed such that subjects were required to produce common responses to members of classes that were not previously associated with one another (e.g., sexual images and members of equivalence class 2, red). Differences in the fluency of performances across both blocks of the final test were sensitive to subjects' relational and conditioning histories. That is, subjects produced significantly more correct responses during block 1 of the test compared to block 2. Such findings lay the foundation for the development*

*of functionally understood behavioral tests and provide a functional-analytic
model of the widely used Implicit Association Test.*

The theoretical and technological basis for the development of
functionally understood implicit behavioral tests has been available for
the past 15 years, although this research has not been capitalized on until
recently (see Roche, Ruiz, O'Riordan, & Hand, 2005 for a chapter-length
review). Specifically, several behavioral studies have employed the concept
of stimulus equivalence to develop implicit tests of behavioral history. By
*implicit*, we mean that the contingencies that control responding in the
test are not verbally discriminable by the subject (i.e., they are outside
conscious awareness). Rather, the contingencies lie largely outside the
experimental preparation, which serves only as a current context to bring
pre-experimentally established behavior to bear. Such a test allows the
researcher to glean specific information about a subject's history without
running the risk of producing experimental demand, social desirability,
or forms of counter-control through the use of explicit questionnaires or
interview methods.

The exciting possibility of developing implicit behavioral tests based
on the concept of stimulus equivalence stems from the crucial finding
that subjects' personal and social histories interfere with the formation of
specific equivalence relations in the laboratory. In a seminal study, Watt,
Keenan, Barnes and Cairns (1991) used a simple stimulus equivalence
paradigm to take advantage of the fact that people in Northern Ireland
often respond to each other's names as indicative of religious background.
Their study employed stimuli representative of Catholic and Protestant
names and symbols and involved training subjects to relate them in a
manner inconsistent with their social histories. During equivalence testing,
all of the English subjects correctly matched the Catholic names with the
Protestant symbols, but 12 of the 19 Northern Irish subjects chose a novel
Protestant name in the presence of the Protestant symbols, thereby failing
to form equivalence relations. These findings strongly suggested that the
social contingencies operating in Northern Ireland interfered with the
establishment of equivalence relations in the laboratory. More specifically,
the equivalence test required Northern Irish subjects to juxtapose names
and symbols in a manner that was countercultural for this group of
subjects. Thus, it would appear that a derived relations paradigm may be
used successfully to assess the social knowledge of participants without
alerting them to the nature of the task.

The stimulus equivalence-based approach has also been employed to
discriminate anxious from nonanxious patients (Leslie, Tierney, Robinson,
Keenan, Watt, & Barnes, 1993), develop a diagnostic tool to identify children
who have been sexually abused (McGlinchey, Keenan, & Dillenburger, 2000;
see also Keenan, McGlinchey, Fairhurst, & Dillenberger, 2000), and to identify
child sex offenders as a distinct social group within a larger population
of non–sex offenders (see Roche et al., 2005). Other researchers have used
the equivalence paradigm to assess subjects' attitudes toward themselves
(Barnes, Lawlor, Smeets, & Roche, 1995; Merwin & Wilson, 2005), toward
sexually explicit stimuli (Grey & Barnes, 1996), as well as attitudes of North
Americans toward Middle Easterners (see Dixon, Dymond, Rehfeldt, Roche, &
Zlomke, 2003).

Interestingly, the foregoing equivalence-based approach to implicit testing appears to be functionally similar to the eponymous Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998). The IAT is a psychometric-style test used to determine what its creators call unconscious bias (e.g., regarding race, age, gender). Specifically, a subject responds to a series of items that may be classified into one of four categories. Two of the categories usually represent a concept (e.g., flowers and insects), while two further categories usually represent a value-laden attribute (e.g., pleasant and unpleasant). Subjects are asked to respond rapidly with a right-hand key press to items representing one concept and one attribute (e.g., insects and pleasant), and with a left-hand key press to items from the remaining two categories (e.g., flowers and unpleasant). Subjects then perform a second task in which the requirements are switched (e.g., flowers and pleasant stimuli share a response and insects and unpleasant stimuli share a response). The IAT records the latencies and accuracies of responses to these two tasks. These measures are interpreted in terms of association strengths. That is, it is assumed that subjects respond more rapidly and accurately when the concept and attribute sharing the same response are pre-experimentally strongly associated (e.g., flowers and pleasant) than when they are weakly associated (e.g., insects and pleasant). In contrast to the vast number of explicit attitude measures, the IAT is thought capable of overcoming experimental demand characteristics and to make it difficult for subjects to fake responses in comparison with explicit tests such as questionnaires (Banse, Seise, & Zerbes, 2001; Kim, 2003).

Despite numerous replications attesting to the reliability of the IAT as a measure of unconscious cognitions, there is a great deal of uncertainty regarding its core processes (e.g., Rothermund & Wentura, 2004). Specifically, the claim that the test measures unconscious (implicit) bias is made in the absence of empirical data (Karpsinki & Hilton, 2001; Steffens & Plewe, 2001). Furthermore, while the original creators of the test make claims regarding its validity, others have called it into question (e.g., De Houwer, 2001, 2006). In spite of a lack of clarity over what exactly the IAT measures and what researchers mean by the term "implicit" (De Houwer, 2006), psychologists continue to apply the IAT from within a poorly understood social-cognitive paradigm (see Fiedler, Messner, & Bluemke, 2006). After an extensive review of the literature, Fiedler et al. (2006) reported the distinct absence of a testable model underlying the IAT. The current research, therefore, served two purposes. The first was to examine a competing contingency approach to implicit testing for social and personal history. The second was to model the IAT test format and to show that it can be understood in terms of a competing contingencies account.

A detailed overview of the current competing contingencies approach to implicit testing was provided by Roche et al. (2005) and was based on a preliminary empirical demonstration of the model developed by Roche, Ruiz and Hand (2003). Roche et al. (2005) argued that the IAT is a measure of subjects' fluency with the relevant verbal categories and their degree of experience at juxtaposing members of those verbal categories (i.e., the extent of contextual control over the categorization of the relevant social/ verbal stimuli). More specifically, the four verbal categories employed in a typical IAT (two concepts and two attributes) are conceived as equivalence classes containing everyday words and objects. It is argued that higher-order equivalence relations (see Wulfert, Greenway, & Dougher, 1994) or relations

between equivalence relations (see Stewart, Barnes-Holmes, Roche, & Smeets, 2002) often obtain in the natural environment. For instance, for a black racist individual the verbal classes *White* and *Bad* might participate in a further higher-order equivalence relation that we may call "things I don't like." The IAT works by measuring the ease with which a common response function (e.g., press the left key) can be established for two or more members of this higher-order equivalence relation compared with members of different and unrelated equivalence relations (e.g., *White* and *Good*). From this perspective, the IAT functions as a subtle or implicit test for derived relations (or for the purpose of a convenient acronym; an Implicit Relational Test; IRT).

To test the foregoing idea, 15 subjects were first exposed to a respondent conditioning procedure in which each of two nonsense syllables, printed in blue and red font, was paired with a sexual or aversive visual image, respectively. Subjects were then exposed to an equivalence training procedure that led to the formation of two three-member equivalence relations, each containing one of the conditioned stimuli as A stimuli. Subjects were then exposed to an equivalence-based IAT-type test consisting of sexual and aversive images and all members of the trained equivalence relations (presented in black font). In one block of the test, subjects were required to produce a common response to members of the same derived equivalence relations, while in another block of the test, they were required to produce a common response to two members of distinct derived equivalence relations. It was expected that subjects would produce more correct responses when common response functions were established for stimuli from the same equivalence relation, compared with when these response functions were established for stimuli from different equivalence relations.

## Method

### Subjects

Fifteen (6 male and 9 female) subjects aged 18–62 years, all acquaintances of the female experimenter, participated in the current study. Subjects were informed that they would be participating in a four-phase word-association test, which would take approximately 1 hr to complete.

### Apparatus

All four phases of the experiment were presented to subjects on an Apple iMac 400 MHz computer with a 15-in. monitor. Stimulus presentations were controlled by using the software package Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993), which also recorded all responses. Two colored abstract shapes and 12 photographic images taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1999) were employed during the conditioning phase. The photographs used as sexual images were classified in the IAPS under the headings "Romance," "Erotic couple, and "Couple." The photographs used as aversive images were classified under the headings "Roaches," "Attack dog," "Disabled," "Electric chair," "Distressed fem," and "Attack." The images used corresponded to the slide numbers 4599, 4601, 4606, 4608, 4609, 4623, 1274, 1525, 3300, 6020, 6311, and 6510. Finally, six nonsense

syllables were employed as stimuli during the equivalence phase. These were Ler, Cug, Mau, Vek, Paf, and Rog. These will be referred to by using the alphanumerics A1, B1, C1, A2, B2, and C2.

## Procedure

*General experimental sequence.* The current experiment consisted of three phases. Phases 1 through 3 were presented consecutively on the computer. Each phase was completed one at a time, and subjects were instructed to contact the experimenter at the end of each phase. The experimenter then initiated the next phase manually. Subjects sat comfortably at a standard computer desk and viewed the computer screen at a distance of approximately 70 cm and at eye level. Phase 1 consisted of word-picture association training and lasted approximately 10 min. Phase 2 consisted of equivalence training and testing during which subjects used a computer mouse to make choices between comparison stimuli on each of a series of baseline conditional discrimination tasks and on probes for derived relations during the equivalence test. Phase 3 consisted of the competing contingencies test and took approximately 10 min to complete.

*Phase 1.* In Phase 1, subjects were exposed to a word-picture association-training procedure in which a respondent conditioning preparation was used. That is, two arbitrary nonsense syllables (A1: Ler, A2: Vek) were paired with sexual and aversive photographic images, respectively. The nonsense syllable paired with the sexual images was blue, while that paired with the aversive images was red. For 5 of the subjects these color associations were reversed (i.e., Ler was red, Vek was blue), but for clarity we will refer only to the original color association configuration.

Subjects were presented with the following instructions on-screen after being seated in front of the computer:

In a moment some words and images will appear on this screen

Your task is to look at these items carefully and to remember what you see

IT IS VERY IMPORTANT THAT YOU CONTINUE TO WATCH THE SCREEN AT ALL TIMES

After each picture has been presented you will be required to press the space bar on the computer to continue. Please make sure you know where the space bar is before you begin.

REMEMBER – IT IS VERY IMPORTANT THAT YOU PLAY CLOSE ATTENTION TO WHAT IS HAPPENING ON THE COMPUTER SCREEN.

If you have any questions please ask them now.

When you are ready please click the mouse button.

All conditioning trials were presented on the computer screen against a black background. A trial began with the presentation of one of the two nonsense syllables appearing in the center of the screen for a period of 2 s followed by an interval of 1 s wherein the screen remained blank. After the 1

s interval the relevant picture appeared in the center of the screen for 4 s. One second after the onset of the image, the nonsense syllable was re-presented in the top left of the screen for the remainder of the trial (i.e., 3 s). In effect, the word-picture association phase employed both a trace and a simultaneous conditioning procedure. At the end of each trial the phrase "Press the space bar" appeared in the center of the screen in 20-point Times font and remained until the subject pressed the space bar. The space bar press functioned as an observing response that initiated the subsequent trial.

There were 10 conditioning trials for each of the two word-picture associations, with no more than two consecutive exposures to either association. Trials were separated by an intertrial interval of 15 s.

*Phase 2.* On completion of the word-picture association-training phase, subjects were immediately exposed to Phase 2, which consisted of equivalence training and testing. Training was designed to lead to the formation of two three-member equivalence relations, each containing one of the two nonsense syllables used during Phase 1 as A stimuli, and two novel nonsense syllables.

There were two baseline conditional discrimination-training tasks, composed of four matching-to-sample tasks. Training was conducted by using a linear training method (i.e., A1-B1, B1-C1 and A2-B2, B2-C2). Prior to training, subjects were presented with brief instructions requesting them to use the computer mouse to click on the comparison stimulus they believed to be correct. Tasks were presented in a quasi-random order in blocks of 16 trials (i.e., four times each with no more than two successive exposures to any task). Subjects matched the comparison stimuli (e.g., Cug, Paf) to the sample (e.g., Ler, Vek) by clicking on the comparison of their choice using the computer mouse and cursor. All choices were followed by corrective feedback delivered by the computer (i.e., the word Correct or Wrong appearing in the center of the screen in red 20-point font for 1.5 s). Feedback informed subjects as to whether their choice was correct or incorrect. Subjects were exposed to successive blocks of training until they responded correctly on 15 of the 16 tasks in a single block (i.e., 93.75% correct). The training tasks were as follows: A1-B1 (B2), A2-B1 (B2), B1-C1 (C2), and B2-C2 (B1), where the stimuli in parentheses indicate incorrect choices.

After reaching criterion during training, subjects were exposed immediately to a block of 16 testing tasks presented in quasi-random order (i.e., four exposures to each of the four tasks, with no more than two successive exposures to any task). The testing tasks were A1-C1 (C2), A2-C2 (C1), C1-A1 (A2), and C2-A2 (A1), where the stimuli in parentheses indicate incorrect choices. Subjects did not receive corrective feedback during this testing period. The testing proceeded, without a break, in blocks of 16 trials until the subject reached the 93.75% correct response criterion in a single block or until 16 blocks had been administered. During testing, subjects were expected to match A1 to C1, C1 to A1, A2 to C2, and C2 to A2, thereby demonstrating stimulus equivalence (see Barnes, 1994; Fields, Adams, Verhave, & Newman, 1990; Sidman, 1986).

Subjects were required to pass the testing phase in order to proceed to Phase 3. All 15 subjects in the current study met the criterion in both the training and testing stages of Phase 2.

*Phase 3.* In Phase 3, subjects were exposed to a discrimination test consisting of sexual and aversive images, and all members of the trained equivalence relations (presented in black font) as stimuli. Recall that in Phase

1, A1 (presented in blue) was paired with sexual images and A2 (presented in red) was paired with aversive images. This phase relied on the well-established transfer-of-functions effect (see Barnes, 1994) to ensure that all remaining class members acquired the relevant color functions. While the derived transfer of color functions to the B and C stimuli was not explicitly tested, it was fundamental to observing a difference in response fluencies across the testing blocks during this phase.

Prior to exposure to this phase, subjects were presented with the following on-screen instructions based on those used by Greenwald et al. (1998).

> In a moment some items will appear on this screen. Your task is to first look at the item and then press either the Z key on the left or the M key button on the right of the keyboard in front of you. Look now to make sure you know where they are. Use the labels at the top of the screen to help you decide which key to press.
>
> Keep in mind:
>
>> Keep your index fingers on the left and right buttons to enable rapid response.
>>
>> Two labels at the top will tell you which words go with each button
>>
>> Each word has a correct classification.
>>
>> Please try to go fast.
>>
>> Expect to make a few mistakes because of going fast. That's OK.
>
> If you have any questions please ask the experimenter now. Click any key when ready to start

Further instructions presented on the top left and right of the screen during all tasks varied, depending on the task block being presented (see Figure 1). The same instructions remained on-screen for the entire duration of a task block and changed only at the beginning of the second task block.

| Press left for<br>Blue and Sexual    Press right for<br>Red and Aversive | Press left for<br>Red and Sexual    Press right for<br>Blue and Aversive |
|---|---|
| Cug | Cug |
| Press left for<br>Blue and Sexual    Press right for<br>Red and Aversive | Press left for<br>Red and Sexual    Press right for<br>Blue and Aversive |
| Paf | Paf |

*Figure 1*. Examples of tasks presented to subjects during the Implicit Relational Test in Phase 3.

Subjects were exposed to a total of 180 trials presented in two task blocks (i.e., 90 trials in each block). The order in which these blocks were presented was randomized across subjects. Each of the blocks consisted of four task types, each of which involved the presentations of one of the following stimulus types in the center of the computer screen: sexual images, aversive images, class 1 stimuli (A1, B1, or C1), and class 2 stimuli (A2, B2, or C2). These four tasks were presented once each in a random order in a block of four trials. There were 22 successive presentations of these four-trial blocks (i.e., 88 trials) followed by two of the four task types chosen randomly by the computer software (i.e., 88 + 2 = 90).

The frequency of presentation of A1, B1, C1, A2, B2, and C2 was equal across the first 88 tasks in each block of 90. The same sexual and aversive images employed during phase 1 were employed during phase 3. Images were chosen randomly by the computer software on relevant trials.

During the relationally consistent task block, subjects were required to make a left-hand key press (i.e., press the Z key on the computer keyboard) upon the presentation of blue and sexual stimuli and to make a right-hand key press (i.e., press the M key on the computer keyboard) upon the presentation of red and aversive stimuli (see Figure 1). In contrast, during the relationally inconsistent task block, red and sexual images shared a left-hand key press and blue and aversive images shared a right-hand key press (see Figure 1). During phase 3, however, all nonsense syllables were presented in black, and so any color functions elicited by the B and C stimuli were derived by virtue of the transfer-of-functions effect (i.e., B1 and C1 should have derived "blue" functions, and B2 and C2 should have derived "red" functions).

Subjects' responses were recorded in terms of both accuracy and latency. However, in contrast to Greenwald's (1998) method of recoding response latencies above 3,000 ms as 3,000 ms, trials in the current study were limited to 3,000 ms duration. This difference was intended to circumvent the problem of devising and negotiating arbitrary statistical procedures designed to extract a hypothetical process from the data set. In effect, subjects were prevented from responding outside the 3,000 ms time frame by the cessation of the trial and the presentation of the subsequent trial. A failure to respond within the 3,000 ms response window was recorded as an incorrect response, and the response latency was recorded as 3,000 ms. Response times were recorded from the trial onset to the first emitted response on the computer keyboard, regardless of whether the response was correct.

## Results

All 15 subjects passed both equivalence training and testing during Phase 2 (see Table 1). The criterion for passing training and testing was set at 15 correct responses in a block of 16 trials. Three subjects passed training on the first block of 16 trials, while 1 subject required 10 blocks to pass training. The mean number of blocks of training required to reach criterion was 4.67. Seven subjects passed the equivalence test on the first block. One subject required 12 blocks of testing to reach criterion. The mean number of testing blocks required to reach criterion was 2.93.

For the purpose of data analysis, response times and accuracies were left in their raw state and not transformed in any way.

Table 1
*Correct Responses for Each Training and Testing Block During Phase 2*

| Subject | Number of Blocks | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1: Train | 15/16 | 16/16 | | | | | | | | | | |
| 1: Test | 16/16 | | | | | | | | | | | |
| 2: Train | 9/16 | 16/16 | | | | | | | | | | |
| 2: Test | 16/16 | | | | | | | | | | | |
| 3: Train | 12/16 | 14/16 | 14/16 | 16/16 | | | | | | | | |
| 3: Test | 16/16 | | | | | | | | | | | |
| 4: Train | 12/16 | 7/16 | 8/16 | 12/16 | 9/16 | 12/16 | 12/16 | 13/16 | 16/16 | | | |
| 4: Test | 16/16 | | | | | | | | | | | |
| 5: Train | 10/16 | 13/16 | 16/16 | | | | | | | | | |
| 5: Test | 16/16 | | | | | | | | | | | |
| 6: Train | 12/16 | 14/16 | 14/16 | 16/16 | | | | | | | | |
| 6: Test | 11/16 | 14/16 | 16/16 | | | | | | | | | |
| 7: Train | 8/16 | 7/16 | 11/16 | 14/16 | 16/16 | | | | | | | |
| 7: Test | 14/16 | 16/16 | | | | | | | | | | |
| 8: Train | 8/16 | 7/16 | 12/16 | 9/16 | 6/16 | 9/16 | 12/16 | 13/16 | 14/16 | 16/16 | | |
| 8: Test | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 0/16 | 3/16 | 16/16 |
| 9: Train | 9/16 | 10/16 | 9/16 | 10/16 | 13/16 | 13/16 | 14/16 | 14/16 | 16/16 | | | |
| 9: Test | 14/16 | 16/16 | | | | | | | | | | |
| 10: Train | 10/16 | 13/16 | 14/16 | 16/16 | | | | | | | | |
| 10: Test | 8/16 | 9/16 | 7/16 | 11/16 | 6/16 | 3/16 | 11/16 | 13/16 | 16/16 | | | |
| 11: Train | 16/16 | | | | | | | | | | | |
| 11: Test | 16/16 | | | | | | | | | | | |
| 12: Train | 16/16 | | | | | | | | | | | |
| 12: Test | 16/16 | | | | | | | | | | | |
| 13: Train | 16/16 | | | | | | | | | | | |
| 13: Test | 14/16 | 14/16 | 16/16 | | | | | | | | | |
| 14: Train | 11/16 | 11/16 | 13/16 | 14/16 | 14/16 | 16/16 | | | | | | |
| 14: Test | 14/16 | 14/16 | 16/16 | | | | | | | | | |
| 15: Train | 4/16 | 7/16 | 12/16 | 4/16 | 5/16 | 5/16 | 11/16 | 11/16 | 16/16 | | | |
| 15: Test | 12/16 | 12/16 | 16/16 | | | | | | | | | |

In Phase 3 all subjects successfully completed the required 180-trial discrimination test. Subjects' total number of correct responses was calculated for both relationally consistent (M = 76.5) and relationally inconsistent (M = 54.6) tasks (see Figure 2). There was considerable variance observed across subjects' total number of correct responses on the relationally consistent and relationally inconsistent tasks.
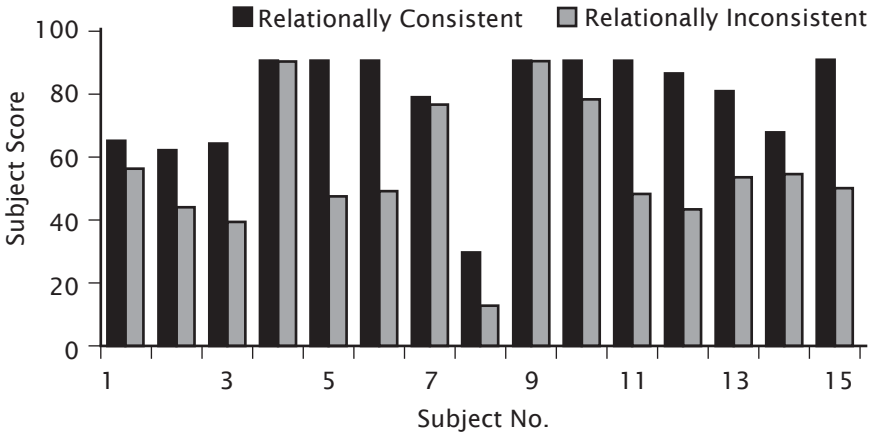


*Figure 2*. Individual subject response accuracies for the Implicit Relational Test in Phase 3.

Subject 8 scored well below chance levels on both the consistent (29) and the inconsistent (12) tasks. Seven additional subjects (S2, S3, S5, S6, S11, S12, and S15) also scored at or below chance levels on the inconsistent tasks. However, all but 1 of the 15 subjects (i.e., S8) scored above chance levels on the consistent tasks, with the lowest of these scores being 60 of a total of 90. It is important that 13 of the 15 subjects responded with greater accuracy on the relationally consistent task block.

An inferential statistical analysis showed that the total correct responses differed significantly across the test blocks in Phase 3 ($t$ = 5.129, $df$ = 14, $p$ = 0.000). The magnitude of the difference across test blocks in terms of the variance in scores within test blocks (i.e., the eta-squared statistic) was 0.652, indicating a large effect size (see Cohen, 1994). Response latencies, however, did not differ significantly between the consistent (M = 1.128 s) and inconsistent (M = 1.176 s) task blocks ($t$ = 0.099, $df$ = 14, $p$ = 0.75). The effect size was also small (eta-squared statistic = 0.001).

## Discussion

The current Implicit Relations Test was clearly sensitive to the laboratory conditioning and relational histories of the subjects. More specifically, both the color-images associations and the equivalence class membership configurations were determined arbitrarily by the experimenters. Nevertheless, response accuracy, in a test that required subjects to make common responses to two stimulus items, was reliably enhanced or retarded by this laboratory-controlled history. Thus, the effect demonstrated here was clearly under experimental control. This effect can be explained only by the congruent and competing contingencies brought to bear during the relationally consistent

and relationally inconsistent bocks of the test. Such a finding suggests that this procedure can be used to assess stimulus associations and equivalence class membership of stimuli in a subject's preexperimental history. Moreover, the procedure is sufficiently subtle that it is doubtful that subjects were aware of the purpose of the critical test (Phase 3). Indeed, none of the subjects was able to report an awareness of the purpose of the test. In other words, subjects were not able to tact the conditioning and relational contingencies controlling their final test performance.

These findings would appear to provide a clear account of the IAT effect in terms of a parsimonious framework without the need to appeal to hypothetical mediating constructs such as attitudes and unconscious cognitions (see Greenwald et al., 1998). This Implicit Relational Test model of the IAT effect employed laboratory-generated stimuli *and* laboratory-created stimulus functions. Thus, this study represents a first step in constructing a testable model of the core processes involved in a very popular psychological test that has previously been difficult to explain (see Fiedler et al., 2006). Our results support the view proposed by Roche et al. (2005) that in the IAT, personal and verbal/cultural histories interfere with responding to relations that juxtapose culturally relevant stimuli in a countercultural fashion.

One previous study, to our knowledge, has been published that attempted to generate an account of the IAT in terms of respondent processes and the juxtaposition of unrelated stimulus classes in the IAT test format. Specifically, Mitchell, Anderson, & Lovibond (2003) taught a group of subjects the "meanings" of four nonwords. Two of these meanings were affectively positive, and two were affectively negative. The researchers found evidence for the transfer of affect in an IAT. That is, the nonwords given pleasant meanings in training were more easily categorized with pleasant than with unpleasant personality characteristics, compared with nonwords given unpleasant meanings.

Respondent processes no doubt play an important role in the development of stimulus classes whose structure can be measured with tests such as that used here. However, the current research extends this view to include the process of derived relational responding. The inclusion of this process in a model of the IAT increases its ecological validity by directly addressing the issue of verbal categories or semantic relations and their widely accepted, though unspecified role in the IAT (see Greenwald et al., 1998).

One important feature of the current data is the clear visibility of individual subject effects in Phase 3. In other words, the test effect was established for most members of the subject population. Specifically, 13 of the 15 subjects responded with greater accuracy on the relationally consistent task block. Moreover, 8 of the subjects responded at or below chance levels on the inconsistent task block, while 3 further subjects responded only marginally above chance level on this block. Thus, a clear and replicated effect was observed for most subjects in the experiment.

No effect was observed for response times with the current test method. However, it is important to understand that this outcome was expected, given that we did not employ an IAT-style statistical methodology. More specifically, the IAT adopts various scoring algorithms (Greenwald, Nosek, & Banaji, 2003; see also Devos and Banaji, 2005) that mathematically derive psychometric scores from an otherwise clear response time and accuracy (i.e., fluency) pattern. In particular, Greenwald and his collaborators employ a response

correction technique that produces a hybrid IAT score combining response time and response accuracy in ways not functionally understood. The procedure involves providing corrective feedback after incorrect responses and requiring a correct response to terminate each trial. Thus, response times reflect the time taken to respond in the first instance *plus* the additional time taken to emit the second altered response. In effect, the IAT delivers a time penalty for those trials on which an incorrect response is emitted, thereby ensuring longer average response times for those tasks that the subject finds most difficult (i.e., relationally inconsistent). It is important to understand, therefore, that what the IAT reports as a response time is more accurately viewed as a procedurally forced and mathematically derived response accuracy and latency hybrid measure and not a direct and transparent index of either. The use of such statistical scoring techniques (which are constantly under revision by Greenwald and colleagues) limit the extent to which the controlling features of a subject's performance can be observed and makes a functional analysis of the behavior under analysis ever more difficult. Indeed, the measurement scales and scoring systems employed in any scientific investigation function as rules that bring the scientist's behavior under the control of aspects of the behavior of interest (Johnston & Pennypacker, 1993). Thus, the analytic methodology limits the response variability of the scientist. To this extent, the radical behaviorist prefers the inductive method (Sidman, 1960) and freedom from the constraints of theory and deduction to as great an extent as possible (Skinner, 1950).

Another reason why our purely descriptive approach to data analysis failed to yield response time differentials relates to our reluctance to employ arbitrary post hoc response time truncation techniques as used in the IAT scoring technique. More specific, the IAT is presumed to measure error rates across task types more effectively when subjects are under time constraints to respond on individual trials. Thus, the IAT involves instructing subjects to respond as quickly as possible on each trial. Nevertheless, no time limit actually is imposed on trials and all response times are recorded, regardless of duration. This strategy allows the researcher to consider their data spread statistically post hoc and to consider outliers differently from study to study, depending on the aims of the research. Typically, IAT researchers recode all response times greater than 3,000 ms to 3,000 ms and all response times less than 300 ms to 300 ms. However, allowing response times of up to 10,000 ms may also be permissible (see Devos & Banaji, 2005; Greenwald et al., 2003). The current study did not employ any such truncation techniques but relied instead on a forced fluency procedure in which speed of responding was enforced on a trial-to-trial basis by a finite response window. It is perhaps because of this behaviorally controlled time constraint that we observed response accuracy differentials where the IAT would fail to see them (i.e., the IAT is based on the aforementioned response accuracy/time hybrid measure rather than accuracy alone). Furthermore, we failed to observe response time differentials where the IAT would typically find them. Thus, the current study illustrates the benefits of behavioral control over the relevant contingencies that produce response differentials across the two task blocks (i.e., relationally consistent and relationally inconsistent) of the IAT.

Even if response time differentials can be generated by using our derived relations model, it remains the case that behavioral measures do not typically emphasize response latencies (but see Bentall, Dickins, & Fox, 1993; Spencer

& Chase, 1996; Steele & Hayes, 1991; Wulfert & Hayes, 1988). This is because response latency is subject to a wide range of interpretations. In particular, behavior analysts are cautious of reaction times measures, as they may be used mistakenly as an explanatory mechanism or as evidence of mediating cognitive processes (e.g., attitudes; see Johnston & Pennypacker, 1993; see also O'Hora, Roche, Barnes-Holmes, & Smeets , 2002; Roche, Linehan, Ward, Dymond & Rehfeldt, 2004). Thus, the current study emphasized accuracy over response time as a measure of the acquisition of the operant discrimination tasks presented in the Implicit Relations Test.

In conclusion, the current research demonstrates that it is possible to use a competing contingencies approach to generate a functionally understood implicit-style test of behavioral history. Moreover, this study has established that it is sufficient to establish a limited number of conditioned response functions and an appropriate network of derived relations, using a range of arbitrary stimuli, to create a laboratory-controlled IAT effect. Thus, it is now clear that IAT effects may in principle be generated in the absence of "unconscious bias" and attitudes, at least as traditionally conceived. Of course, behaviorally speaking, an attitude may be conceived as a network of derived and explicitly reinforced stimulus relations according to which the functions of events are transformed (Grey & Barnes, 1996). To this extent the current implicit relations test may indeed be said to measure attitudes thus defined. What is clear from the current study, however, is that the IAT may reveal no more than the organization of an individual's conditioning and verbal history through the use of a competing contingencies test. Therefore, the current derived relations model builds on the seminal work conducted by Watt et al. (1991) in laying the foundation for the development of functionally understood implicit tests that can rival the IAT as reliable and valid measures of behavioral history.

## References

BANSE, R., SEISE, J., & ZERBES, N. (2001). Implicit attitudes toward homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145–160.

BARNES, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record, 44*, 91–124.

BARNES, D., LAWLOR, H., SMEETS, P., & ROCHE, B. (1996). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record, 46*, 87–107.

BENTALL, R. P., DICKINS, D. W., & FOX, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology, 46B*, 187–214.

COHEN, J. (1994). The Earth is round (*p* < .05). *American Psychologist, 49*, 997–1003.

COHEN J. D., MACWHINNEY B., FLATT M., & PROVOST J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers, 25*, 257–271.

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. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage Publishers.

DEVOS, T., & BANAJI, M. R. (2005). American = white? *Journal of Personality and Social Psychology, 88*, 447–466.

DIXON, M. R., DYMOND, S., REHFELDT, R. A., ROCHE, B., & ZLOMKE, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues, 12*, 129–147.

FIEDLER, K. A., MESSNER, C., & BLUEMKE, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology, 17*, 74–147.

FIELDS, L., ADAMS, B. J., VERHAVE, T., & NEWMAN, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53*, 345–358.

GREENWALD, A. G., MCGHEE, D. E., & SCHWARTZ, J. L. (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: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

GREY, I., & BARNES, D. (1996). Stimulus equivalence and attitudes. *The Psychological Record, 46*, 243–270.

JOHNSTON, J. M., & PENNYPACKER, H. S. (1993). *Readings for strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

KARPINSKI, A., & HILTON, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774–788.

KEENAN, M., MCGLINCHEY, A., FAIRHURST, C., & DILLENBERGER, K. (2000). Accuracy of disclosure and contextual control in child abuse: developing procedures within the stimulus equivalence paradigm. *Behavior and Social Issues, 10*, 1–17.

KIM, D. (2003). Voluntary controllability of the Implicit Association test (IAT). *Social Psychology Quarterly, 66*, 83–96.

LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1999). *International Affective Picture System: Instruction manual and affective ratings. Technical Report A-4,* The Center for Research in Psychophysiology, University of Florida.

LESLIE, J. C., TIERNEY, K. J., ROBINSON, C. P., KEENAN, M., WATT, A., & BARNES, D. (1993). Differences between clinically anxious and non-anxious subjects in a stimulus equivalence training task involving threat words. *The Psychological Record, 43*, 153–161.

MCGLINCHEY, A., KEENAN, M., & DILLENBURGER, K. (2000). Outline for the development of a screening procedure for children who have been sexually abused. *Research on Social Work Practice, 10*, 721–747.

MERWIN, R. M., & WILSON, K. G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *The Psychological Record, 55*, 561–575.

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

O'HORA, D., ROCHE, B., BARNES-HOLMES, D., & SMEETS, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52*, 51–75.

ROCHE, B., LINEHAN, C., WARD, T., DYMOND, S., REHELDT, R. (2004). The unfolding of the relational operant: A real-time analysis using electroencephalography and reaction time measures. *International Journal of Psychology and Psychological Therapy, 4*, 1–17.

ROCHE. B., RUIZ, M. R., & HAND, K. (2003). *An experimental analysis of social discrimination using relational frame theory.* Paper presented at the Annual Conference of the Association for Behavior Analysis, San Francisco, May 23–27.

ROCHE, B., RUIZ, M., O'RIORDAN, M., & HAND, K. (2005). A relational frame approach to the psychological assessment of sex offenders. In M. Taylor & E. Quayle (Eds.), *Viewing child pornography on the Internet: Understanding the offence, managing the offender, and helping the victims* (pp. 109–125). Dorset, UK: Russell House Publishing.

ROTHERMUND, K., & WENTURA, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.

SIDMAN, M. (1960). *Tactics of scientific research.* New York: Basic Books.

SIDMAN, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. E. Zeiler (Eds.), *Analysis and integration of behavioural units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

SKINNER, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*, 193–216.

SPENCER, T. J., & CHASE, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659.

STEELE, D., & HAYES, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56*, 519–555.

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

STEWART, I., BARNES-HOLMES, D., ROCHE, B., & SMEETS, P. (2002). A functional-analytic model of analogy: A relational frame analysis. *Journal of the Experimental Analysis of Behavior, 78*, 375–396.

WATT, A., KEENAN, M., BARNES, D., & CAIRNS, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record, 41*, 33–50.

WULFERT, E., GREENWAY, D., & DOUGHER, M. (1994). Third-order equivalence classes. *The Psychological Record, 44*, 411–439.

WULFERT, E., & HAYES, S. C. (1988). The transfer of conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125–144.