Psychol Rec (2014) 64:475–485
DOI 10.1007/s40732-014-0033-z

ORIGINAL ARTICLE

A Derived Transformation of Evaluative Preferences Using Implicit Association Tests

Micah Amd · Dermot Barnes-Holmes

Published online: 31 May 2014

Abstract In the current experiment, multiple implicit association tests (IATs) were employed to examine the transformation of emotional functions across stimuli that have been related along a comparatively valenced ("happier" to "unhappier") dimension. Ten human participants were exposed to a matching-to-sample (MTS) procedure where they were trained to select the more positively valenced ("happier") stimulus in the presence of a yellow contextual cue or otherwise to select the more negatively valenced ("unhappier") stimulus in the presence of a red cue. Next, the cues were employed to establish the relations A > B, B > C and C > D where ">" indicates "happier than." Following tests for mutual and combinatorial entailment, participants underwent two single-category IATs, where the A-D and B-C stimulus pairs were alternatively paired with happy and unhappy words. Our results indicate that individuals who demonstrated evidence of mutual and combinatorial entailed relations paired Stimulus A (more so than Stimulus D) and Stimulus B (more so than Stimulus C), with happy words more fluently then with unhappy words.

Keywords Implicit association test · Comparative relations · Valence

A wide range of studies have reported the derived transfer/transformation of functions in accordance with multiple stimulus relations such as Same/Opposite, Before/After, and More/Less relations (Dymond and Whelan 2010; Munnelly et al. 2010; Dymond and Barnes 1995; Munnelly et al. 2013). The most basic example of this process involves the transfer of functions through a stimulus equivalence class; in an example of a recent study involving equivalence classes, Amd et al. (2013) trained participants to match three arbitrary stimuli—A-B, and B-C—and then tested for the formation of the predicted equivalence relations—B-A, C-B, A-C, and C-A. Prior to the establishment of the equivalence relations, Stimulus A had been paired with positively valenced (appetitive) pictures such that A came to elicit neurophysiological responses indicative of positive affect. The derived transfer of functions through the equivalence class was then demonstrated when Stimulus C was also shown to elicit similar emotive responses, although C had not been directly paired with Stimulus A or emotive stimuli previously.

The concept of derived *transformation* (rather than *transfer*) of functions is typically reserved for describing those instances in which the functions of stimuli emerge based on relations other than equivalence, such as *comparative* relations (see Hayes et al. 2001). Briefly, comparative relations involve relations between stimuli that differ in magnitude along a prespecified functional dimension. For example, in a seminal study by Dougher et al. (2007), comparative relations were established using three arbitrary stimuli, such that A was established as less than B, and B was established as less than C. The B stimulus was then paired with the delivery of a moderate electric shock, and A was paired with the delivery of a single shock that was half the strength of the B stimulus shock. Subsequent testing showed that participants produced levels of physiological arousal that were predictably lower for A relative to B, but when C was presented, arousal levels were considerably higher (than for B). In this case, the arousal functions of the A stimulus and particularly C (because it had never been paired with any shock at all) had been transformed in accordance with the derived comparative relation. Or more colloquially, if B means moderate shock and A

M. Amd · D. Barnes-Holmes
National University of Ireland, Maynooth, Maynooth, Ireland

M. Amd (✉)
Department of Psychology, National University of Ireland,
Maynooth, Maynooth, Co. Kildare, Ireland
e-mail: kudos.ma@gmail.com

means minimal shock, and C is more than B, then C means greater shock.

In assessing the emotional valence of a stimulus, participants may be asked to rate the stimulus using Likert-type scales and other self-report measures (see De Houwer et al. 2001, for a review), although the use of physiological and indirect performance-based measures are becoming more common (Amd et al. 2013; Hinojosa et al. 2010; Smith et al. 2004; Schupp et al. 2003). In regard to the latter, some researchers have suggested that it may be wise to determine if specific performance-based measures also capture the derived transformation of functions. One reason for suggesting the use of such measures is that they appear to be less susceptible to experimenter-expectation effects and individual response biases than are self-reports. For example, O'Toole et al. (2007) argued that participants may generate self-rules during a derived transfer study based on how they think the experimenter wants them to behave, and then respond accordingly. As such, the observation of a derived transformation of functions effect may reflect control by variables that extend beyond those that were the target of the experiment (e.g., compliance with socially mediated contingencies). Or more informally, the derived transformation might arise, in part, out of a participant's desire to please the experimenter. Additionally, self-reported ratings may come to serve as contextual cues for self-rule generation given that utilizing forced-choice ratings can inadvertently establish valence functions for the stimulus rated. In other words, merely asking a person to rate a stimulus with a socially derived construct ("moderately happy") might establish an equivalence class between the stimulus and the evaluative properties of the rating itself, potentially transforming the functions of the stimulus rated prior to and during the experimental procedure (Wallaert et al. 2010; Lane and Critchfield 1996).

In an effort to circumvent such effects, O'Toole et al. (2007) suggested that it may be wise to explore the use of performance-based measures that are less prone to deliberate control by participants. To test this suggestion, O'Toole et al. employed a measure that has been used widely across many domains in psychological science, the implicit association test, or IAT (Greenwald et al. 1998; see Hughes et al. 2011, for a detailed treatment of the relationship between self-report and performance-based measures such as the IAT, from both cognitive and behavioral perspectives). Briefly, the IAT requires individuals to categorize stimuli, quickly and accurately, into pairs. In some blocks of trials the stimulus pairs are deemed to be congruent with participants' preexperimental histories, and in other blocks of trials the stimulus pairs are deemed to be incongruent with their histories.

In the first part of the O'Toole et al. (2007) study, participants were asked to press one button if a picture of a baby or a romantic couple was presented on the computer screen and to press another button if a picture of a snake or a spider was presented. These trials were defined as congruent because babies and romantic couples were deemed to be positively valenced, whereas spiders and snakes were deemed to be negatively valenced. During incongruent trials, participants were asked to press one button if a picture of baby or a picture of a snake was presented, and to press a second button if a picture of a romantic couple or a spider was presented; in effect, the valences of the stimuli to be categorized together were deliberately mixed. As expected, participants responded with greater fluency (i.e., higher accuracy and lower response latencies) during congruent relative to incongruent blocks of trials.

The critical part of the O'Toole et al. (2007) study involved establishing four separate equivalence classes for each of the types of pictorial stimuli employed in the previously outlined IAT. Thus, one class consisted of nonsense syllables that were related directly or indirectly to pictures of babies, another class consisted of syllables that were related to pictures of romantic couples, a third class consisted of syllables that were related to pictures of snakes, and the fourth class consisted of syllables that related to pictures of spiders. Participants were then asked to complete the IAT for a second time, but instead of being presented with the actual pictures of the babies, couples, snakes, and spiders, the nonsense syllables from the four classes were presented. The basic prediction was that the positive and negative valences of the four different types of pictures should transfer via equivalence relations to the nonsense syllables and thus produce an IAT effect similar to the one observed when the actual pictures were presented. The results of the experiment supported this prediction.

As noted previously, the emergence of novel behaviours based on the transfer of functions via equivalence relations constitutes only one way in which derived stimulus relations may produce such effects. At the time of writing, however, there were no published studies documenting *transformation* effects using the IAT or any similar performance-based measures. This was the primary purpose of the current research.

In the current experiment, participants were presented with emotionally valenced images (of faces) that differed along a comparative dimension of valence, from happy to neutral to unhappy. By a *comparative dimension* we refer to the relative, qualitative nature of valence functions for a given stimulus in comparison to another stimulus within the experimental context. In the current experiment, the image of a smiling person (S1) may be deemed as more positively valenced when contrasted with a second image of the same person but with an unhappy expression (S2). For the sake of explanatory clarity, we may than posit that Stimulus S1 is more positively valenced, or happier, than Stimulus S2 (and

S2 is more negatively valenced, or unhappier, than S1), or S1 > S2 and S2 < S1.[1]

In the current study, participants were required to choose one of two images in the presence of a particular background color (contextual cue). For example, given a yellow background color, choosing a happy face over an unhappy or neutral face or choosing a neutral face over an unhappy face was reinforced. Given the color red, choosing an unhappy face over a neutral or happy face or a neutral face over a happy face was reinforced. The aim, therefore, was to establish the color yellow as functionally analogous to "happier" and the color red as analogous to "unhappier." These two cues were then used to establish comparative relations among four stimuli (A, B, C, & D) that did not differ physically from each other in terms of emotional valence. For example, for some participants, given the color yellow and Stimuli A-B, selecting A (rather than B) was reinforced. The aim here, therefore, was to establish the arbitrarily applicable relations, "A happier than B" and "B unhappier than A." Additional such relations were trained (i.e., "B happier than C" and "C happier than D") and then tested (e.g., "C unhappier than B"; "B happier than D"; "D unhappier than A").

In the final phase, two IATs were implemented. In the congruent blocks of trials of the first IAT, participants were required to press one key whenever Stimulus A or a positively valenced word was presented, and to press another key if Stimulus D or a negatively valenced word was presented. In the incongruent blocks, categorizing A with negatively valenced and D with positively valenced words was required. The second IAT was similar to the first except that Stimulus A was substituted with Stimulus B and Stimulus D was substituted with Stimulus C. The two IATs thus sought to determine if the relative "happy" and "unhappy" functions of the stimuli within the four-member network had been transformed so that Stimulus A (more so than Stimulus D) and Stimulus B (more so than Stimulus C) would be categorized more fluently with positively valenced (happy) words. It is important to note that the use of two separate IATs allowed us to test the comparative properties of the relational network. Specifically, if only one IAT was employed (e.g., assessing the A-D relation) it could be argued that only a frame of opposite was being measured (i.e., because A and D lie at the opposite ends of a continuum). However, because a second IAT was used (testing the B-C relation), the comparative nature of the network can be assessed.

Method

Participants

Ten undergraduate psychology students from the National University of Ireland, Maynooth, (mode=19 years of age) were recruited via a random sampling method. All participants were right-handed, native-English speakers with normal or corrected-to-normal vision. Participants performed the experiment in a 330 cm×250 cm×220 cm whitewashed room within a temperature range of 17±1.5 °C. The experiment took approximately 60 min to complete per participant.

Apparatus

Tasks were designed on E-Prime 2.0 with parameters adapted from previous relational responding experiments. Images of individuals classified under happy (V+), unhappy (V-), and neutral (V_N) categories were taken from the freely available Radboud Faces Database (Langner et al. 2010). The ratings of the faces were acquired from the same dataset. A total of 30 images consisting of 10 individuals (five male, five female) with three expressions each (happy, neutral, unhappy) were used during the training and testing of contextual cues (Phase 1). During Phase 1, each trial depicted two images from either category such as happy-unhappy, neutral-happy, neutral-unhappy; for sake of clarity, all happy (positively valenced) faces will be referred to as V+, all unhappy (negatively valenced) faces will be referred to as V−, and faces with neutral expressions will be referred to as V_N. Stimulus A and Stimulus D (in Phase 2) consisted of the images of a "grey" and "orange" face, respectively, against a black background. Stimulus B and Stimulus C consisted of the images of an "orange" and "grey" Cabbage Patch doll, respectively, against a white background (see Fig. 1). The hues were acquired through filtering using Pixlr Editing photo-editing software. All participants were instructed to face a 42.7 cm Dell computer screen from a length of 69± 5 cm, at a viewing angle of −29±1°, and had to respond on a standard QWERTY keyboard. All experiments were programmed and presented in E-Prime 2.0, a program designed for psychological experiments.

Procedure

Phase 1

The first phase consisted of a 60-trial training block (or blocks) followed by a block of 120 test trials, which consisted six trial types, each presented 20 times (see Table 1 for a schematic

[1] In principle, of course, an individual can be described as happy on Day 1 (e.g., because the sun is shining) and even happier on Day 2 (e.g., because the sun is shining and he or she wins the lottery). Thus, happiness (and unhappiness) may be conceptualized as occurring along a comparative dimension. It is also important to note that although *happy* versus *unhappy* may be conceptualized as participating in a frame of opposition; in the current study, happy, unhappy, *and* neutral face stimuli were employed, and thus the frame of comparison is required to accommodate relational responding among these three stimuli (see Dymond and Barnes 1995, for a detailed conceptual and empirical treatment of this issue).

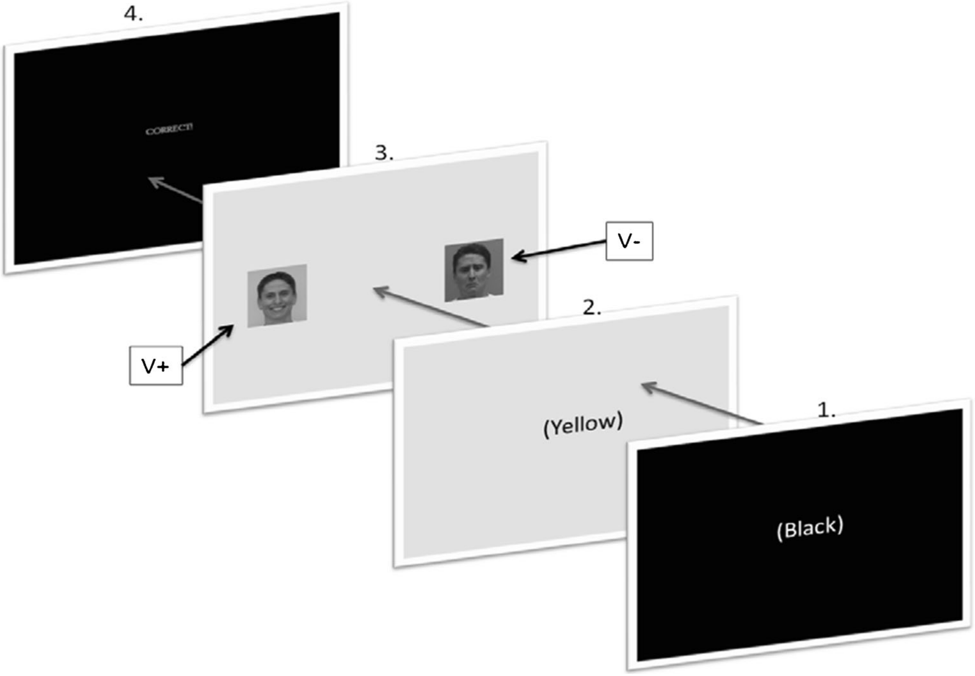

Fig. 1 Stimuli used for establishing the four member network A-B-C-D

representation of the trial types). The experiment commenced with the following instructions presented on screen:

> Welcome to the first part of the experiment! In what will follow, you will be presented with two images on screen in either a RED or YELLOW background. You will have to select one of the images to continue. One of the images will be "correct," depending on the BACKGROUND COLOUR. To select the image on the left, please press "a." To select the image on the right, please press "l." You will be provided with corrective feedback as you progress. Please take your time—it is important to respond accurately rather than quickly. If you have any questions, please ask the experimenter. Otherwise, you may press any key to begin …

Table 1 Trial-types and comparisons presented during testing in Phase 1

Trial type	Comparisons*	Background**	Correct response
1	V+ , V−	Yellow	V+
2	V+ , V_N	Yellow	V+
3	V− , V_N	Yellow	V_N
4	V+ , V−	Red	V-
5	V− , V_N	Red	V-
6	V+ , V_N	Red	V_N

*(V+) refers to images depicting smiling/laughing individuals; (V−) refers to images of those same individuals frowning/crying; (V_N) refers to those individuals with emotionally neutral expressions

**Refers to the background color used during training (i.e., with corrective feedback)

Upon pressing a key, the training block initiated with a blank, black screen for 500 ms, after which the background of the screen changed into either a yellow or red hue for 1,000 ms. Next, two differentially valenced images (i.e., V+/V−, V+/V_N, or V-/V_N) appeared on the left and right sides of the computer screen for 5,000 ms, during which time participants had to select one of the images by pressing the appropriate key on the keyboard (i.e., pressing *a* on the keyboard to select the image on the left or *l* to select the image on the right). Responses made on other keys were not recorded and had no programmed function. Participants 1 through 5 (referred to as P1, P2, P3, P4, and P5) underwent training in the presence of the yellow cue only, where selecting the relatively happier face in the presence of the yellow background produced the word "Correct!" in a green font on a black screen for 1,500 ms; selecting the relatively unhappier face produced the word "Wrong" in a red font for 1,500 ms. Participants 6 through 10 (P6, P7, P8, P9, and P10) underwent training in the presence of a red background only, where selecting the relatively happier face in the presence of the red background produced the word "Wrong" in a red font (on a black screen), whereas selecting the unhappier image produced the word "Correct!" in a green font for 1,500 ms (see Fig. 2).

If no selection was made within 5,000 ms of the two images being presented, the statement "Too slow …" appeared in a grey font on a black screen for 1,500 ms. An intertrial interval (ITI) consisting of an empty black screen presented for 500 ms followed the feedback message, after which the subsequent trial began. Participants underwent training until they made 20 correct responses consecutively before the end of a single 60-trial training block; failure to do

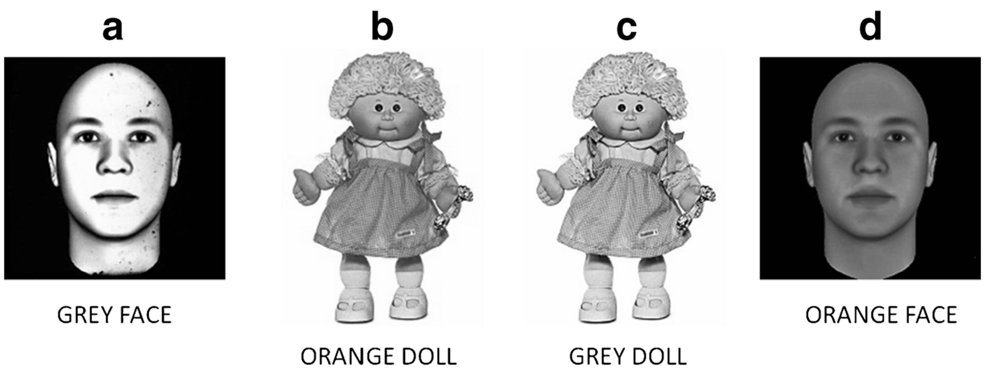

Fig. 2 Illustration of training trials in Phase 1. A blank, black screen was displayed for 500 ms (1), followed by a change in the background color of the screen to either red or yellow for 1,000 ms (2). Two comparisons were presented against the colored background until the participant produced an acceptable response—in the example illustrated, a happy face (V+) and an unhappy face (V−) are presented against a yellow background, and selecting the V+ comparison was followed by the word "Correct!" (4)

so resulted in reexposure to the training block. If after three reexposures a participant was still unable to reach the response accuracy criterion, the computer was programmed to continue on to Phase 2. All participants achieved the desired response accuracy criterion within the first training block except for P1, P4, P6, and P7, who achieved the criterion on the second iteration of the training block. Upon completion of the training trials, participants commenced testing with the following instructions presented on screen:

> You will stop receiving feedback in the following trials. Please take your time and respond accurately. When ready, press any key to begin …

Upon pressing a key, participants had to select among the comparisons in the presence of both yellow and red backgrounds an equal number of times. Stimulus presentation parameters were kept similar to the training phase, with the major difference being the removal of the feedback screens and the inclusion of a longer ITI (2,000 ms) in the form of an empty black screen. The completion criterion for the test block was a response accuracy criterion of at least 18 correct responses out of 20 for each trial type. Trials were presented in a quasirandom sequence, such that no two trial types which were the same were presented consecutively. Each of the six trial types differed in terms of the two stimuli and/or the contextual cue (color background) that was presented (see Table 1), such that all possible elements were used to test relational responding (e.g., Trial Type 1: happy face and an unhappy face presented on a yellow background; Trial Type 5: unhappy face with a neutral face on a red background, etc.).

Phase 2

Training trials were similar to those used in Phase 1, except that the comparison stimuli consisted of Stimuli A, B, C, and D (pictures of orange and grey faces and dolls) rather than images of happy, unhappy, and neutral faces. Phase 2 initiated with the following instructions:

> Welcome to the second part of the experiment! Similar to before, you will be presented with two images on screen in either a RED or YELLOW background. You will have to select one of the images to continue. One of the images will be "correct," depending on the BACKGROUND COLOR. To select the image on the left, please press "a." To select the image on the right, please press "l." You will be provided with corrective feedback as you progress. Please take your time and respond accurately. If you have any questions, please ask the experimenter. Otherwise, you may press any key to begin …

Participants P1, P2, and P3 were trained to select Stimulus A (from Comparisons A-B), Stimulus B (from B-C) and Stimulus C (from C-D) in the presence of a yellow background only, effectively establishing the relations A > B > C > D, where ">" indicates "happier (than)." P4, P5, and P6 were trained to select Stimulus B (from A-B), Stimulus C (from B-C), and Stimulus D (from C-D) in the presence of a red background, establishing the relations D < C < B < A, where "<" indicates "unhappier (than)." P7, P8, P9, and P10 underwent training in both yellow and red backgrounds for individual trial types. Similar to Phase 1, participants underwent training until they could produce 20 consecutively correct responses before moving on to the testing phase.

It should be noted that P7 and P8 were exposed to A > B and A > C trial types during training (see Table 2 for trial types of Phase 2), and thus the derived relation between B and C remains unspecified (see Vitale et al. 2008). As such, P7 and P8 served as control participants for which no specific IAT effects could be predicted. In contrast, P9 and P10 were trained to establish a network (B > C, C > D, B < A) which was relationally coherent with the network established for P1, P2, P3, P4, P5, and P6 (A > B, B > C, C > D), and thus these

Table 2 Schematic representation of training and testing trial types for mutual entailment (ME) and combinatorial entailment (CE) in Phase 2

Participant	Training trial types	ME test trial types	CE test trial types
P1, P2, P3	C_y*(A>B**, B>C, C>D)	C_R(B<A, C<B, D<C)	C_y(A>C, B>D, A>D); C_R(C<A, D<B, D<A)
P4, P5, P6	C_R(B<A, C<B, D<C)	C_Y(A>B, B>C, C>D)	C_y(A>C, B>D, A>D); C_R(C<A, D<B, D<A)
P7, P8	C_Y(A>B); C_R(C<B, D<C)	C_R(B<A); C_y(B>C, C>D)	C_y(A>C, B>D, A>D); C_R(C<A, D<B, D<A)
P9, P10	C_R(D<C, C<B); C_Y(A>B)	C_Y(B>C, C>D); C_R(B<A)	C_y(A>C, B>D, A>D); C_R (C<A, D<B, D<A)

*Subscript refers to the color (context) that the comparisons in parentheses were presented. Specifically, C_Y refers to the background color *yellow*, whereas C_R refers to the color *red*

** Illustrates the direction of the trained/tested relation. For example, A > B indicates that Stimulus A is *happier* or more positively valenced than B

two participants were predicted to demonstrate the hypothesized IAT effects.

After training, all participants underwent tests for what will be described as mutual entailment and combinatorial entailment (see Hayes et al. 2001), which were conducted without corrective feedback. Simply, mutual entailment assesses for derived bidirectionality between related stimuli (if A is trained as happier than B, then B is derived to be unhappier than A) whereas combinatorial entailment refers to the emergence of a third relation upon the combination of two or more previously trained relations (if A is trained as happier than B and B is trained as happier than C, then responding to A being happier than C would be indicative of combinatorial entailment).

An example of a test for mutual entailment was as follows: Given that P1 was trained to select A (from Comparisons A-B) in the presence of a yellow background, choosing B in the presence of a red background was defined as a correct mutually entailed response (i.e., if A is happier than B, then B is unhappier than A). Critically, participants were presented with pairs of stimuli in the presence of a background color not used during training during the test trials. In testing for combinatorial entailment, participants were required to select from stimulus pairs that were at least one-node apart (i.e. A-C, B-D and A-D) and had never been presented together previously during training. For P1, an example of a test for combinatorial entailment involved presenting the comparisons A-C in either a red or yellow background. Given that A was trained as happier than B (A > B) and B as happier than C (B > C), then A is happier than C, or A > C. Other tests for combinatorial entailment included selecting from B-D and A-D comparisons.

Tests for mutual and combinatorial entailment consisted of blocks of nine trials (three mutual entailment and six combinatorial-entailment trial types), which were presented three times in succession for a total of 27 trials. The response accuracy criterion required that participants produce seven correct (out of nine responses) for mutual entailment trial types, and 15 correct (out of 18 responses) for combinatorial-entailment trial types. Participants who failed to reach the accuracy criteria (7/9 and 15/18) were labelled as a “FAIL” group; those who reached the criteria were labelled the “PASS” group. The sequence of training and test trial types presented to participants is shown in Table 2.

Phase 3

In the final phase, participants were presented with two separate IATs. In general terms, the first IAT (IAT1) required that participants categorize the grey and orange faces (Stimulus A and Stimulus D), with words deemed to be consistent with their newly derived valences for some blocks of trials (i.e., the A stimulus with happy words and the D stimulus with unhappy words), and for other blocks of trials to categorize the faces in a manner that was deemed to be inconsistent with their derived valences (i.e., the A stimulus with unhappy words and the D stimulus with happy words). The same parameters applied to IAT2, with the exception that participants were asked to categorize the orange and grey Cabbage Patch dolls (the B and C stimuli) rather than the orange and grey faces.

For IAT1, the concept categories were titled GREY FACE and ORANGE FACE, with the stimulus members for the concept categories comprising the A and D stimuli, respectively. The attribute categories were titled HAPPY and UNHAPPY. The stimuli that were defined as happy for the purposes of the current study were the words *overjoyed, satisfied, grand,* and *glad*; the stimuli defined as unhappy were the words *miserable, unhappy, depressed,* and *weary.* For IAT2, the concept categories were titled ORANGE DOLL and GREY DOLL, with the stimulus items for the concepts comprising of Stimulus B and Stimulus C, respectively, while the attribute categories (and their resultant members) remained the same as for IAT1. At the beginning of the both IATs, participants were presented with the following instructions:

> You will be presented with a set of words or images to classify into groups using the "a" and "l" keys on the keyboard. Classify the items quickly while making as few errors as possible. You may make some errors at first, and that is okay. You will get better as you progress. Please ask the experimenter if you have any questions. Otherwise, press any key to begin …

After a key press on the keyboard, the words GREY FACE and ORANGE FACE appeared on the left and right sides of the top half of the screen in size 14 Times Roman font, in white against a black background, along with the presentation of either Stimulus A or D in the centre of the screen for the first trial of Block 1 (B1). The block consisted of 20 trials, and for each trial participants had to press the *a* key (which is on the left of a QWERTY keyboard) if Stimulus A (a grey face) appeared in the centre of the screen. If Stimulus D (an orange face) appeared, pressing the *l* key (which is on the right of the keyboard) was required. Correct responses simply cleared the screen for 150 ms, and the next trial was then presented. If an incorrect response was made, a red X appeared on screen below the face stimulus, prompting the participant to emit the "correct" response, which was then followed by the 150 ms blank screen before the onset of the next trial.

Block 2 (B2) was similar to B1 except the labels GREY FACE and ORANGE FACE were replaced with the words HAPPY and UNHAPPY, and one of the happy or unhappy words appeared in the middle of the screen on each trial. Participants were thus required to press the *a* key if a happy word appeared and to press the *l* key if an unhappy word appeared.

For the third block (B3), both concept and attribute categories were presented on screen so that the words GREY FACE and HAPPY were presented together on the left side, while ORANGE FACE and UNHAPPY were presented together on the right side of the screen. Participants were presented with 20 trials and were required to respond in the same manner as across B1 and B2 (pressing the *a* key for Stimulus A and happy words and pressing the *l* key for Stimulus D and unhappy words). Block 4 (B4) was a continuation of B3, but 40 rather than 20 trials were presented.

Block 5 (B5) was similar to Block 1, except the left–right positioning of the labels GREY FACE and ORANGE FACE were reversed. Participants were thus required to reverse their response pattern established across previous blocks (i.e., press *l* in the presence of Stimulus A, and press *a* in the presence of Stimulus D).

Blocks 6 (B6) and 7 (B7) were similar to B3 and B4, except the reversed positioning of the two "face" labels employed in Block 5 was used. Participants thus had to press the *a* key if Stimulus D or a happy word appeared on screen and the *l* key if Stimulus A or an unhappy word was presented. IAT2 was a reiteration of the same procedure, albeit with concept categories ORANGE DOLL and GREY DOLL (instead of ORANGE FACE and GREY FACE) and with Stimulus B and Stimulus C replacing Stimulus A and Stimulus B (see Fig. 3 for an illustration of individual IAT trials). Response latencies and errors from Blocks 3, 4, 6, and 7 were collected and analysed in accordance with the recommendations of Greenwald et al. (2003).

Results

Phase 1

All participants achieved the criterion of 20 consecutively correct responses in the training phase. Participants P2, P3, P5, P8, P9, and P10 achieved the response accuracy criterion in the first iteration of the 60-trial training block; the remaining participants—P1, P4, P6, and P7—achieved the criterion in the second iteration of the training block (see Table 3). Testing included six trial types that were presented in blocks of six trials. A total 120 test trials were presented (i.e., 20 blocks), during which participants selected between happy (V+), unhappy (V-), and emotionally neutral (V_N) faces in the presence of yellow or red backgrounds. Participants P2, P3, P5, P8, P9, and P10 matched the response accuracy criterion, producing at least 18 out of 20 correct responses for each trial type. Averaged response accuracy for P4 and P7 exceeded 80 %, but they failed to maintain the criterion for each of the trial types. The response accuracy for P1 and P6 fell well below criterion, approximating 60 % correct (see Table 4).

Phase 2

All participants met the training criterion of 20 consecutively correct responses in fewer than 60 trials (see Table 5 for details). In the test phase, participants were exposed to nine trial types (three for mutual entailment and six for combinatorial entailment) presented in three successive blocks of nine trials. Participants P2, P3, P5, P7, P9, and P10 met/exceeded the specified response accuracy criteria (7/9 and 15/18 correct responses for mutual and combinatorial entailment, respectively); the remaining participants—P1, P4, P6, and P8—failed to do so (see Table 5).

Phase 3

The two IATs were implemented upon completion of Phases 1 and 2 for all participants. Response latency and accuracy data were analysed in accordance with Greenwald and colleagues' (2003) recommended protocols. Only data from Blocks B3,

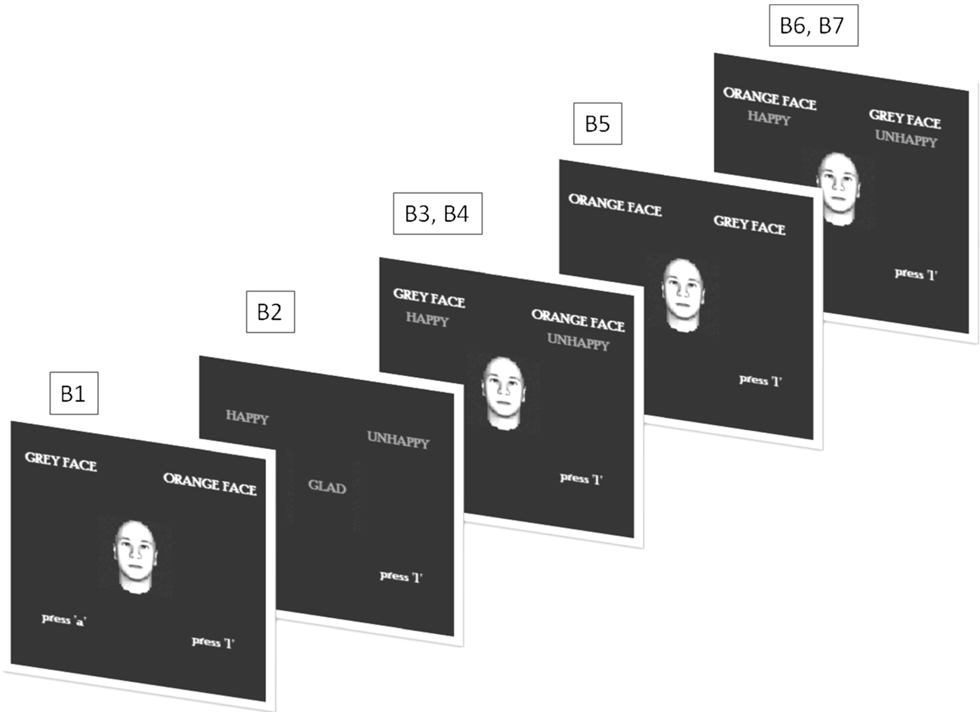

Fig. 3 Illustration of the different stages of IAT1. In Block 1 (B1), either Stimulus A or Stimulus D were presented and had to be paired with the words GREY FACE or ORANGE FACE. In B2, happy or unhappy words were presented and had to be paired with the words HAPPY or UNHAPPY. In B3 and B4, the words GREY FACE and HAPPY were presented on one side of the screen, while the words ORANGE FACE and UNHAPPY were presented on the other side of the screen. Participants were presented with Stimuli A, D, happy words, or unhappy words, which they had to pair with the concepts presented on either side of the screen. B5 was similar to B1, with the location of the words GREY FACE and ORANGE FACE switched. B6 and B7 were similar to B3 and B4, with the exception of the alternate positioning of the words GREY FACE and ORANGE FACE. In IAT2, the words GREY FACE and ORANGE FACE were replaced with GREY DOLL and ORANGE DOLL, with Stimuli B and C replacing Stimuli A and D, respectively (not shown)

B4, B6, and B7 were used in the final analysis; no participant produced response latencies >10,000 ms or <300 ms, and thus all trials were used for analysis. Four mean latencies were computed from all the correct response latencies in each of the four blocks. To adjust for response errors, latencies for incorrect responses were replaced with the block means (of the correct responses only) increased by an increment of 600 ms. For example, if the response latency for an incorrect response in B3 was700 ms and the mean response latency for all correct responses produced during B3 was 900 ms (μ), the latency reported for the incorrect response would be (μ + 600)=1,500 ms. Next, one pooled standard deviation (*SD*1) was computed for all trials in B3 and B6, and another (*SD*2) was computed for all trials in B4 and B7.

Table 3 Response accuracies (by individual participant) for training trials over Phase 1

Participant	Train 1 Errors*	Train 2 Errors
1	27	3
2	8	*n/a*
3	11	*n/a*
4	31	11
5	12	*n/a*
6	17	3
7	21	9
8	9	*n/a*
9	12	*n/a*
10	11	*n/a*

* Gives the total number of incorrect responses made prior to emitting 20 consecutively correct responses. Train 1 indicates the number of errors recorded for the first iteration of the 60-trial training block. Train 2 indicates the number of errors observed for the second iteration of the 60-trial training block

All adjusted latency values from each of the four trial blocks were then computed into four means, B3μ, B4μ, B6μ, and B7μ. Next, the adjusted means for B3 and B6 were divided by the standard deviation calculated from across these two blocks (B3μ + B4μ/*SD*1). Similarly, the latencies for Blocks 4 and 7 were divided by the standard deviation calculated from the latter two blocks (B4μ + B7μ/*SD*2). The two values were averaged to give a so-called *D*(600)-score (Karpinski and Steinman 2006; Greenwald et al. 2003). Briefly, the *D* score controls for interindividual variability in response speeds and association strength while providing an index of preference for one pair of IAT categories more so than another (Cai et al. 2002).

Table 4 Response accuracies per trial type per individual participant for test trials during Phase 1

Participant	TT1*	TT2	TT3	TT4	TT5	TT6	Test Accuracy (%)**
1	12	15	10	12	12	10	0.59
2	18	20	18	19	18	18	0.93
3	18	20	18	18	18	18	0.90
4	18	18	15	16	16	14	0.81
5	20	20	18	20	18	18	0.95
6	12	14	10	13	15	10	0.62
7	16	14	15	18	15	15	0.78
8	18	18	18	18	18	18	0.90
9	18	18	18	18	19	18	0.91
10	19	18	19	18	20	19	0.94

*Refers to the trial types illustrated in Table 1—the number of correct responses per trial type (out of 20 total responses) are presented

**Response accuracy averaged over 120 trials expressed in decimal percentiles

For current purposes, a positive *D* score for IAT1 and IAT2 indicates that Stimulus A more so than Stimulus D (for IAT 1), and Stimulus B more so than Stimulus C (for IAT 2), was more fluently paired with happy words than with unhappy words. Conversely, negative *D* scores indicate Stimulus A was more fluently paired with unhappy than with happy words and Stimulus D was more fluently paired with happy rather than unhappy words (IAT1), and Stimulus B was more fluently paired with unhappy words than was Stimulus C. For the PASS group, which had showed the emergence of derived comparative relations, it was predicted that *D* scores for both IAT1 and IAT2 would be positive. Consistent with this prediction, Participants 2, 3, 5, 7, 9, and 10 showed positive *D* scores for both IATs. Participants who were classified as in the FAIL group failed to show this consistent pattern of positive *D* scores across both IATs (see Table 6). If we assume that the *random* probability of producing either a negative or positive *D* score in any single IAT is .5, the probability of producing two positive (or two negative) scores across both IATs would be .25. Given that six participants produced positive *D* scores across both IATs, the probability of all six participants acquiring these results by chance alone would be .25/6=.042. That is, the probability of all six participants in the PASS group producing positive *D* scores across both IATs was significantly less than .05. Although the sample size is small, the probability of attaining our results by chance remains unlikely (<.05) and provides support for the experimental hypothesis.

Table 5 Response accuracies for training and testing trials over Phase 2

Participants	Train 1 Errors*	Mutual Entailment**	Combinatorial Entailment
1	12	5/9	8/18
2	10	8/9	16/18
3	19	8/9	17/18
4	14	5/9	8/18
5	9	7/9	15/18
6	15	4/9	7/18
7	11	7/9	15/18
8	12	5/9	9/18
9	16	7/9	15/18
10	22	8/9	16/18

*Gives the total number of incorrect responses prior to producing 20 consecutively correct responses over a 60-trial training block

**Gives the number of correct responses (out of total responses) for each test trial type. Critically, the response accuracy (RA) criterion for mutual entailment was eight correct (out of nine responses); the RA criterion for combinatorial entailment was 15 correct (out of 18 responses)

Table 6 Computed difference scores for IAT1 (A vs. D) and IAT2 (B vs. C) per individual participant*

Participants	GROUP**	IAT1 (A vs. D)	IAT2 (B vs. C)
1	FAIL	0.69	−0.37
2	PASS	0.49	0.34
3	PASS	0.62	0.32
4	FAIL	−0.69	−0.61
5	PASS	0.43	0.49
6	FAIL	0.56	0.44
7	PASS	0.43	0.21
8	FAIL	−0.87	0.36
9	PASS	0.79	0.81
10	PASS	0.68	0.91

*Positive *D* scores indicate that Stimulus A was more readily paired with *happy* words than Stimulus D (for IAT1) or that Stimulus B was more readily paired with *happy* words than Stimulus C (for IAT2). Alternatively, a negative *D* score would imply that Stimulus D was more readily paired with *happy* words than Stimulus A (for IAT1) or that Stimulus B was more readily paired with *happy* words than Stimulus C (for IAT2). Participants have been categorized on the basis of their response accuracies during the test trials in Phase 2

**Individuals who met the RA criterion in Phase 2 for both mutual and combinatorial entailment were deemed to be part of the PASS group; all others were parsed under the FAIL group

Discussion

Ten participants underwent MTS procedures and two IATs over three phases to establish and assess for the derived transformation of emotional functions across a four-member comparative relational network. In the first phase, participants had to select between happy, unhappy, and emotionally neutral faces in the presence of either a yellow or red background. In the presence of a yellow background, participants were trained to select the happier (more positively valenced) face whereas in the presence of a red background, participants were trained to select the unhappier (more negatively valenced) face. In essence, the yellow background color served as a contextual cue for selecting the relatively positively valenced stimulus whereas the red background color served as a cue for selecting the relatively negatively valenced stimulus. In the training block of the second phase, Participants 1 to 6 had to select from comparisons A-B, B-C, and C-D in the presence of either a yellow (Participants 1, 2, 3) or red (Participants 4, 5, 6) background during training. Participants 7, 8, 9, and 10 had to select from the comparisons in the presence of both yellow and red backgrounds. In the ensuing test block, mutual and combinatorial entailment were assessed by presenting, without feedback, the A-B, B-C, and C-D comparisons in the presence of a background color not used during training (mutual entailment), along with Comparisons A-C and B-D that were never directly presented with one another at a previous stage (combinatorial entailment). In the final phase of the current study, participants were required to pair the A-D stimuli (IAT1) and the B-C stimuli (IAT2) rapidly (and accurately) with happy and unhappy words. All participants who showed strong evidence of mutual and combinatorial entailment across the test blocks of Phase 2 demonstrated IAT effects indicative of the transformation of so-called valence functions across the four-member relational network. At the time of writing, no published study had employed a performance-based measure, such as the IAT, to assess the *transformation* of valence functions across a comparative relational network, and as such the current findings are unique.

Although the current findings are reasonably compelling, it must be acknowledged that further research is required to develop and refine the current line of inquiry. First, given there were no baseline IAT measurements of stimuli A-B-C-D with happy/unhappy words, clear experimental control over the IAT performances was not provided here. Indeed, ideally, establishing an IAT effect in one direction and then reversing the effect based on derived a transformation of functions would be needed to demonstrate transformation of any so-called “valence” function. On balance, the inherent reversibility of an IAT performance should not be taken for granted (Gregg et al. 2006), and thus any attempt to reverse an established IAT effect may raise issues that extend beyond establishing an appropriate transformation of functions.

Second, given that the network employed in the current study had only four stimuli, the IAT effects seen for Stimuli A and D may have arisen from stimulus–stimulus associations rather than the transformation of functions per se. Consider that for three participants (1, 2, & 3), Stimulus A was always selected, and Stimulus D was never selected in the presence of yellow (happy). It is possible that the A stimulus for these participants may have acquired its “happy” functions via direct pairing with a cue that had previously been established as functionally equivalent to “choose happier.” As such, the effect observed with the IAT may have emerged in part due to the transfer of functions via an equivalence relation rather than a comparative network. On balance, the inclusion of the second IAT assessing the valence functions of the B and C stimuli would be difficult to explain in these terms, given that they were both selected in the presence of yellow an equal number of times. One must recognise, however, that the relationship between B and C is one of mutual, rather than combinatorial, entailment, and this fact limits the current interpretation of the findings in terms of a “full-blown” transformation of functions.

Another possible criticism is that the background colors of red and yellow were used as contextual cues to establish the happier-than versus unhappier-than functions, and then different hues (orange and grey) were used with the faces and dolls. Perhaps, therefore, unspecified or uncontrolled primary stimulus generalization effects among the stimuli could be used to explain the findings of the current study (see Hoon et al. 2008). On balance, this does seem highly unlikely, given the results observed; if hue was a determining factor in the IAT-performances, than the observed results would be A > D, C > B (i.e., if “grey” was deemed inherently more appetitive than “orange”). Conversely, we would have observed D > A, B > C if orange was deemed inherently more appetitive than grey, rather than the results we attained. None of the participants who passed the training and testing phases demonstrated this pattern (see Table 6).

Future research could address these issues by employing baseline as well as postprocedure IATs, establishing a larger relational network (to counter for possible higher conditioning or mediated generalization effects) and by reversing the direction of valence function transformations across multiple networks. In any case, the current findings provide initial support for the argument that it should be possible to assess the derived transformation of valence functions using performance-based measures such as the IAT.

Authors Notes The current study was conducted as part of the doctoral research program undertaken by the first author at the National University of Ireland, Maynooth. The research was initially funded by the John Hume scholarship and subsequently by the Irish Research Council for Science, Engineering and Technology. The author thanks Bryan Roche for the clarification of select conceptual issues.

References

Amd, M., Barnes-Holmes, D., & Ivanoff, J. (2013). A derived transfer of eliciting emotional functions using electroencephalograms as a dependent measure. *Journal of the Experimental Analysis of Behavior, 99*(3), 318–334. doi:10.1002/jeab.19.

Cai, H., Sriram, N., & Greenwald, A. (2002). The implicit association test's *D* measure can minimize a cognitive skill confound: comment on McFarland and Crouch. *Social Cognition, 22*(6), 673–684.

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior, 88*(2), 179–197. doi:10.1901/jeab.2007.45-05.

Dymond, S., & Barnes, D. (1995). A transfer of self-discrimination functions in accordance with the derived stimulus relations of sameness, more-than, and less-than. *Journal of the Experimental Analysis of Behavior, 64*(2), 163–184. doi:10.1901/jeab.1995.64-163.

Dymond, S., & Whelan, R. (2010). Derived relational responding: a comparison of match-to-sample and the relational completion procedure. *Journal of the Experimental Analysis of Behavior, 94*(1), 37–55. doi:10.1901/jeab.2010.94-37.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853–869. doi:10.1037//D033-29O9.127.6.853.

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*(6), 1464–1480. doi:10.1037/0022-3514.74.6.1464.

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*(2), 197–216. doi:10.1037/0022-3514.85.2.197.

Gregg, A. P., Seibt, B., & Banaji, M. B. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*(1), 1–20. doi:10.1037/0022-3514.90.1.1.1.

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

Hoon, A., Dymond, S., Dixon, M. R., & Jackson, J. W. (2008). Contextual control of slot machine gambling: replication and extension. *Journal of Applied Behavior Analysis, 41*, 467–470. doi:10.1901/jaba.2008.41-467.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: propositional and behavioral alternatives. *The Psychological Record, 61*, 465–498.

Hinojosa, J. A., Bertolo, C. M., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: behavioral and neural correlates. *Psychophysiology, 47*(4), 748–757. doi:10.1111/j.1469-8986.2010.00982.x.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16–32. doi:10.1037/0022-3514.91.1.16.

Lane, S., & Critchfield, T. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior, 65*(2), 355–374. doi:10.1901/jeab.1996.65-355.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Knippenberg, A. V. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24*(8), 1377–1388. doi:10.1080/02699930903485076.

Munnelly, A., Dymond, S., & Hinton, E. C. (2010). Relational reasoning with derived comparative relations: A novel model of transitive inference. *Behavioural Processes, 85*(1), 8–17. doi:10.1016/j.beproc.2010.05.007.

Munnelly, A., Freegard, G., & Dymond, S. (2013). Constructing relational sentences: Establishing arbitrarily applicable comparative relations with the relational completion procedure. *The Psychological Record, 63*(4), 751–768. doi:10.11133.j.tpr.2013.63.4.004.

O'Toole, C., Barnes-Holmes, D., & Smyth, S. (2007). A derived transfer of functions and the implicit association test. *Journal of the Experimental Analysis of Behavior, 88*(2), 263–283. doi:10.1901/jeab.2007.76-06.

Schupp, H. T., Markus, J., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science, 14*(1), 7–13. doi:10.1111/1467-9280.01411.

Smith, A. P., Dolan, R. J., & Rugg, M. D. (2004). Event-related potential correlates of the retrieval of emotional and nonemotional context. *Journal of Cognitive Neuroscience, 16*(5), 760–775. doi:10.1162/089892904970816.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitated responding in accordance with the relational frame of comparison: systematic empirical analyses. *The Psychological Record, 58*(3), 365–390.

Wallaert, M., Ward, A., & Mann, T. (2010). Explicit control of implicit responses: Simple directives can alter IAT performance. *Social Psychology, 41*(3), 152–157. doi:10.1027/1864-9335/a000022.