A Derived Transformation of Valence Functions Across Two 8-Member Comparative Relational Networks

Micah Amd[1] · **Bryan Roche**[1]

Abstract The emergence of transitive relations between stimuli that had never been directly paired with one another can be examined through a phenomenon called Transitive Inference (TI). The present experiment explored contextually controlled TI effects in verbally able humans. Specifically, participants were trained in the conditional discriminations A1+B1-, B1+C1-, C1+D1-, D1+E1-, E1+F1-, F1+G1- and G1+H1- in the presence of a cue (Cue 1), followed by tests for mutual and combinatorial entailment in the presence of either Cue 1 or Cue 2. Note that Cue 1 and Cue 2 had been previously established as functionally equivalent to *happier-than* and *unhappier-than* contexts, respectively. Using a performance-based measure of "implicit preferences," we predicted that successfully demonstrating entailment would yield a performance indicating C1 as more positively valenced than F1. Similarly, if participants learned the discriminations A2+B2-, B2+C2-, C2+D2-, D2+E2-, E2+F2-, F2+G2-, G2+H2- in the presence of Cue 2 only, followed by tests for entailment in the presence of both Cues 1 and 2, we predicted that C2 should be responded to as more negatively valenced than F2. Performances across both conditions supported these predictions, furthering the evidence for the claim that emotional valences can be derived through functionally transitive stimulus-stimulus relations.

Keywords Relational frame theory · Transitive inference · Transformation of function · Valence

✉ Micah Amd
micah.amd.eab@hotmail.com

[1] Maynooth University, Maynooth, Leinster, Ireland

One of the ways researchers have explored the ability of humans and nonhumans to produce untaught, emergent, or unreinforced responding is through a phenomenon called Transitive Inference, or TI (Martin and Alsop 2004; Vasconcelos 2008). A simple example of a TI task that may be used with English-speaking humans could involve presenting the statements "Adam is happier than Bob; Bob is happier than Carry; Carry is happier than Dan; Dan is happier than Emur," followed by the question "is Bob happier than Dan?" Replying "yes" would be indicative of a so-called Transitive Inference (TI). Critically, given that Bob's happiness is never directly related to Dan's happiness in the original statements, providing the correct answer requires making an inference based on the "mediating node" (Carry) in the transitive relation. It should be noted that TI has been demonstrated across various nonhumans as well (Vasconcelos 2008). An example of a TI protocol typically used with nonhumans could involve first training successive pairs of overlapping discriminations, such as A+B-, B+C-, C+D- and D+E-, where "+" indicates the reinforced stimulus choice. During subsequent probe trials, a novel or untrained pair, such as B-D, can be presented. If the organism selects B more frequently than D in the absence of differential reinforcement, this may be taken as evidence for TI (Guez and Audley 2013). The fact that nonhumans have produced TI performances suggests that transitive inference per se may not necessarily require advanced verbal reasoning abilities specific to humans.

One approach to complex verbal behavior, however, maintains that there may be certain features of complex TI effects unique to verbally sophisticated humans (Relational Frame Theory, or RFT; Hayes et al. 2001). According to RFT, humans learn to relate stimuli in increasingly complex and subtle ways as they interact with the verbal community from an early age, culminating in the emergence of what has been termed Arbitrarily Applicable Relational Responding (AARR; Berens and Hayes 2007; Hayes et al. 2001). Simply put, AARR refers to relational

responding that can come under the control of arbitrary contextual cues. Once established, AARR may provide a foundation for many performances that involve verbal reasoning, inference and derivation (Munnelly et al. 2010; Steele and Hayes 1991). A critical feature of AARR involves the role of contextual control when relating various stimuli. Traditionally, two broad classes of contextual control have been defined by RFT (Hayes et al. 2001). The first of these describes contextual control over a particular *type* of relational responding, such as that provided by the phrase "is more than," as in "A is more than B." In this case, RFT would deem the phrase "is more than" as a relational context, or "C_{rel}," as it specifies a relation type (of 'more than') between events A and B. Concurrently, a second class of contextual control involves specifying the psychological dimension along which a given stimulus–stimulus relation applies. For example, if presented with the statement "X is more *happy* than Y," the word *happy* specifies the relevant psychological dimension (i.e., levels of happiness) and is deemed a C_{func}. It should be noted that within natural language, both C_{rel} and C_{func} control may be exerted by a single event. For instance, if we rephrase "A is more *happy* than B" into the functionally equivalent "X is *happier* than Y," one could imagine the "happi" and "er" parts of "happier" as representative of C_{func} and C_{rel} control, respectively. When both types of contextual control are provided concurrently (e.g., the word *happier*), the cue may simply be described as a $C_{rel+func}$ cue. Although we employ English words in the previous example, a recent study (Amd and Barnes-Holmes 2014) established such $C_{rel+func}$ control using differentially colored computer backgrounds.

In the Amd and Barnes-Holmes study, the colors yellow and red were established as functionally equivalent to the phrases *happier than* and *unhappier than*, respectively. This was achieved by presenting emotional face pairs (e.g., happy vs. unhappy) in a yellow or red background context, where selecting the happier face in a yellow context, or the unhappier face in a red context, were reinforced. Next, participants were trained in three overlapping discriminations involving arbitrary, emotionally homogenous stimuli (A+B-, B+C-, C+D-) in the yellow context only. Participants were then tested for the emergence of what has been defined as mutual and combinatorial entailed relations where, for example, if A+B- and B+C- were trained in a yellow context, then discriminating B+A- or C+B- in a red context would be a mutually entailed response. Additionally, discriminating C+A- (in a red context) or A+C- (in a yellow context) would be combinatorially entailed responses. It may be noted that tests for entailment are functionally indistinguishable from tests for symmetry and transitivity within the context of TI[1]. What differentiates RFT from traditional TI accounts, however, is the prediction that, following entailment, the functions of the transitively related stimuli can

transform as a direct consequence of the context in which the transitive relations emerged. That is, RFT predicts that the "happiness" functions of the A-B-C-D stimuli would differ along a comparative dimension A>B>C>D, where A would be the happiest and D the least happy. In other words, the functions of the stimuli would *transform* in accordance with the entailed relations (Amd and Barnes-Holmes 2014; Amd et al. 2013; De Almeida and de Rose 2015; Dougher et al. 1994, 2002; Munnelly et al. 2010; Roche and Dymond 2008).

To test for this (derived) transformation of valence functions, Amd and Barnes-Holmes employed two Implicit Association Tests[2] (IATs; Greenwald et al. 1998). In one IAT, participants had to pair the A stimulus with happy words and the D stimulus with unhappy words in some blocks of trials, and then reverse these pairings in other blocks of trials (i.e., A with unhappy and D with happy). Pairing was achieved by successfully producing a simple discrimination response (e.g., press the "d" key) to any two stimuli across distinct trials. The second IAT was similar, except that participants had to pair the B and C stimuli with happy and unhappy words alternatively across intermixed blocks of trials. The results from both IATs were consistent with the authors' prediction, in that participants who had demonstrated mutual and combinatorial entailment paired stimulus A (relative to D) and stimulus B (relative to C) with happy words more readily than vice versa. Critically, Amd and Barnes-Holmes's (2014) was the first study to show a transformation of derived valence functions across a comparative relational network using a performance-based measure of so-called implicit attitudes (but see O'Toole et al. 2007, for a study employing the IAT as a measure of the derived transfer of functions across equivalence classes).

Although the results of Amd and Barnes-Holmes (2014) were promising, there was a key conceptual limitation to the research. Recall that the relational network required responding to only three premise pairs, A+B-, B+C- and C+D-. Hence, any IAT response patterns indicating A as happier than D might have arisen from the fact that A was always reinforced as "happier" and D as "unhappier" during relational training, culminating in what can be described as the end-anchor effect (Delius and Siemann 1998). Hence, the observed A-D effect might not have been "fully" derived. Of course, the same criticism would not apply against the IAT that tested the B-C relation because the selections of B and C were both reinforced (B+C-, C+D-) and nonreinforced (A+B-,

---

[1] For the sake of consistency, the term 'entailment' was employed throughout the remainder of the manuscript.

[2] Procedurally, 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. The underlying assumption is that congruent stimulus pairs (e.g., a picture of a smiling face with the word *happy*) will be paired together more quickly and accurately (i.e., *fluently*) than incongruent pairs (e.g., smiling face with *unhappy*).

B+C-) during training. Critically, however, any transformation effect observed with the B-C relation could be explained by the facts that (i) B had already been reinforced as the correct discrimination when the B-C pair appeared in the presence of the *happier-than* contextual cue, and (ii) B was nodally adjacent to A, the selection of which was always reinforced. Therefore, assessing a B-C relation would involve only mutual, and not combinatorial, entailment. Hence it could be argued that the findings reported by Amd and Barnes-Holmes was not a clear demonstration of a derived transformation of valence (happiness) functions across a comparative relational network involving both mutual and combinatorial entailment. A key purpose of the present study, therefore, was to demonstrate a transformation of functions that involved both types of entailment.

The current research extends the study reported by Amd and Barnes-Holmes (2014) in a number of important ways. First, the present study involved training and testing two eight-member relational networks across two days of experimentation (rather than a single four-member network). On Day 1, participants were trained in seven overlapping pairs of conditional discriminations in the presence of a single contextual cue—Cue 1 (A1+B1-, B1+C1-, C1+D1-, D1+E1-, E1+F1-, F1+G1-, G1+H1-), that is, A1>>H1—whereas on Day 2, they were trained a different set of conditional discriminations in the presence of a different contextual cue—Cue 2 (A2+B2-, B2+C2-, C2+D2-, D2+E2-, E2+F2-, F2+G2-, G2+H2-), that is, A2<<H2. It should be noted that Cue 1 and Cue 2 were previously established as *happier-than* and *unhappier-than* cues, respectively, for select participants.

Second, the present study involved the use of baseline and post network establishment IATs to measure relative changes in the valence functions of the C and F stimuli. Specifically, the stimulus pairs subjected to IATs were C1-F1 (from the A1>>H1 network) and the C2-F2 pair (from the A2<<H2 network). This was an important improvement on Amd and Barnes-Holmes because a single IAT could yield an effect consistent with the predicted transformation of function effect simply by chance (i.e., 50 % odds on each IAT outcome). Taking a baseline IAT measure, however, allowed us to determine if the relational training and testing would produce a *relative* change in the IAT outcome in the predicted direction. In other words, incorporating a baseline versus post assessment allows us to gauge changes in participant performances as a direct consequence of network training and testing.

In summary, the present study involved a within-participant design where IAT performances served as the repeated measure for assessing relative stimulus valences, or 'happiness'. On Day 1 of experimentation, an eight-member comparative network (A1>>H1) was trained in the presence of Cue 1 (established as *happier than*) only and, on Day 2, a second eight-member network (A2<<H2) was trained in the presence of Cue 2 (established as *unhappier than*) only. Subsequent tests for entailment took place in the presence of

both Cue 1 and Cue 2. Baseline and post IATs assessed the relative valence functions of the C1-F1 pair (from the A1>>H1 network) as well as the C2-F2 pair (from the A2<<H2 network), both before and after relational training. The C1-F1 and C2-F2 pairs were selected as they were nodally equidistant from each other and their respective end anchors, mitigating for the possibility of end-anchor and value-transfer effects (cf. Weaver et al. 1997). Note that all discriminations on Day 1 were trained (i.e., reinforced) in the presence of Cue 1 *only*, whereas all discriminations on Day 2 were trained in the presence of Cue 2 *only*. In other words, participants were never exposed to both Cue 1 and Cue 2 within the same conditional discrimination training block; rather, participants were exposed to both cues only during nonreinforced test trials.

Given the foregoing procedure, it was predicted that following demonstration of entailment, the happiness functions of C1 should be *more* than F1 for A1>>H1, whereas the happiness functions of C2 should be *less* than F2 for A2<<H2. That is, for A1>>H1, the predicted transformative effect was C1>$_{happy}$ F1, whereas for A2<<H2, the predicted effect was F2>$_{happy}$ C2.

Method

Participants

Twelve male nonpsychology students from the Building Resources Across Communities (BRAC) University in Dhaka, Bangladesh, were recruited for the present study in accordance with institutional ethical requirements. Two participants withdrew from the experiment at various stages for personal reasons, leaving a final sample of 10 (Mode= 23 years, Mean=23.5 years). Monetary compensation was provided in an amount equivalent of 20 USD to each individual participant in increments of 10 USD per day. All participants were trained and tested individually. All participants completed the study in a well-lit $406 \times 356 \times 305$ cm^3 room at typical room temperature, in which they were instructed to face a 43 cm laptop screen from a length of 27 ± 1 in. The participants were parsed into an Experimental group (P1, P2, P3, P4, P5, P6) and a Control group (P7, P8, P9, P10). Participants were not aware of their group classifications. The duration of the experiment averaged 8 hours per participant over the course of two training and testing days; the break between these two days never exceeded 48 hours.

Apparatus and Materials

All experimental tasks were designed and presented in *E-Prime Professional Studio 2.0* (Schneider et al. 2002). All tasks were presented on a 43 cm high-definition laptop screen. Images of 18 human faces with emotionally neutral

expressions were taken from the Radboud face database (Langner et al. 2010) and are labeled here; X1, X2, A1, A2, B1, B2, C1, C2, D1, D2, E1, E2, F1, F2, G1, G2, H1 and H2 (see Fig. 1b for examples of stimuli used). Stimulus items X1 and X2 (not shown) were used in the practice IAT only. The bottom half of all face stimuli were masked with a black rectangle. All stimuli were matched on the spatial dimensions of their outlying borders (0.9 in. × 1.0 in.) and contrast ratio (495:1 to 505:1 candela/m^2). The cues and exemplars used were designed on Microsoft Paint. Specifically, Cue 1 was a black triangle, Cue 2 was a gray square, Cue 3 was a blue star, and Cue 4 a green hexagon. Note that Cues 3 and Cue 4 were contextual cues for the "control" condition and have not been presented in Fig. 1b. During training and testing for *more-than* and *less-than* functions, 12 exemplars (M1–M6, L1–L6) depicting different quantities of various objects were employed. These included images of greater versus fewer quantities of apples (M1 vs. L1), cars (M2 vs.L2), people (M3 vs.L3), red circles (M4 vs.L4), green lightning bolts (M5 vs.L5), and black lines (M6 vs.L6). During training and testing for *happier-than* and *unhappier-than* functions, 16 "smiley" faces served as exemplars. These ranged from "very happy" (V++1, V++2, V++3, V++4), "happy" (V+1, V+2, V+3, V+4), "unhappy" (V-1, V-2, V-3, V-4) to "very unhappy" (V–1, V–2, V–3, V–4).

Procedure Summary

The experimental procedure involved multiple phases and segments, which are summarized below (see Fig. 1a).

Day 1

1. IAT practice phase—Participants were habituated with the IAT procedure.
2. Phase 1a (*more-than/less-than* training and testing)— Cue 1 and Cue 2 were established as *more-than* and *less-than* cues, respectively, through selection of a larger number of objects reinforced in the presence of Cue 1 and selection of fewer numbers of objects reinforced in the presence of Cue 2. Both the Experimental and Control groups received this training.
3. Phase 1b (*happier-than/unhappier-than* training and testing)—Cue 1 and Cue 2 were established as *happier-than* and *unhappier-than* cues, respectively, through selection of relatively happier faces reinforced in the presence of Cue 1 and selection of relatively unhappier faces reinforced in the presence of Cue 2 *only* for the Experimental group. The Control group underwent similar training and testing but with two novel cues, Cue 3 and Cue 4, established as *happier-than* and *unhappier-than* cues, respectively. Note that the order of Phases 1a and 1b were counterbalanced between participants.

4. Phase 2 (IAT-1a)—Participants underwent an IAT where they had to pair C1 with happy words and F1 with unhappy words, and then switch the pairings. Note that the order of pairing was counterbalanced across participants.
5. Phase 3 (establish A1>>H1)—Participants underwent training and testing in order to establish an eight-member network, A1>B1>C1>D1>E1>F1>G1>H1, where all conditional discriminations (e.g., A1+/B1-; B1+/C1-) were trained in the presence of Cue 1 only, whereas tests for entailment took place in the presence of both Cue 1 and Cue 2.
6. Phase 4 (IAT-1b)—IAT-1a was re-administered for all participants. We predicted for the Experimental group that C1 would be paired with happy words and F1 with unhappy words more readily than had been observed at baseline.

Day 2

7. Phase 5a (more/less probes)—The *more-than/less-than* functions established in Phase 1a/Day 1 were probed for during non-reinforced test trials.
8. Phase 5b (happier/unhappier probes)—The *happier-than/unhappier-than* functions established in Phase 1b/Day 1 were probed for during non-reinforced test trials.
9. Phase 6 (IAT-2a)—Participants underwent a second IAT where they paired C2 with happy words and F2 with unhappy words. Similar to IATs 1a and 1b, the order of pairings were counterbalanced across participants.
10. Phase 7 (establish A2<<H2)—Participants underwent training and testing in order to establish a second eight-member network, A2<B2<C2<D2<E2<F2<G2<H2. In contrast to Phase 3, all conditional discriminations (e.g., A2+/B2-, B2+/C2-) were trained in the presence of Cue 2 only. All tests for entailment took place in the presence of both Cue 1 and Cue 2. See Fig. 3.
11. Phase 8 (IAT-2b) –IAT-2a was re-administered for all participants. We predicted that, for the Experimental group, F2 would be paired with happy words, and C2 with unhappy words, more readily then during baseline.

Day/Condition 1

IAT Practice Phase

A practice IAT was employed to familiarize participants with the IAT procedure before the main study commenced. The practice IAT, along with all other IATs employed for the remainder of the experiment, were single-category IATs similar to the ones employed in Amd and Barnes-Holmes (2014). Participants were required to sort one exemplar per concept

category with four exemplars per attribute category within a set of seven response blocks. At the beginning of the experiment, the participant's age, gender and handedness were noted by the experimenter. Participants were subsequently presented with the instructions to classify presented words/images into groups using the "a" and "l" keys on the keyboard.

In the practice IAT, the labels JIM and SAM appeared on the left and right sides of the top half of the screen in size 14 point Times Roman font in yellow, along with the presentation of either the X1 or X2 stimulus in the center of the screen. Participants were required to press the "a" key when X1 was presented or the "l" key when X2 was presented across 20 trials in the first block. If a response was not emitted within 3,000 ms of stimulus onset, a message stating "Too slow!" appeared near the bottom of the screen followed by presentation of the subsequent trial for 1,000 ms. Incorrect responses were followed by a red "X" on screen and correct responses were followed by the message "Correct!" in a green font on the screen for a period of 1,000 ms. An intertrial interval (ITI) of 150 ms followed the 1,000 ms feedback screen, followed by onset of the next trial (see Gavin et al. 2008, and Ridgeway et al. 2010, for a justification of modifications to the typical IAT procedure).

After 20 trials, the labels JIM and SAM were removed and the labels HAPPY and UNHAPPY appeared on the left and right sides of the screen in a blue font in the second block. Participants were presented with happy words (*overjoyed, glad, satisfied, delighted*) and unhappy words (*miserable, mournful, sad, depressed*) as stimuli and were required to press "a" in the presence of a happy word or to press "l" in the presence of an unhappy word. After another 20 trials, the labels JIM and HAPPY appeared on the left side of the screen and the labels SAM and UNHAPPY appeared on the right side of the screen. Participants were presented with the X1 and X2 stimuli interspersed with happy and unhappy words and were required to press "a" in the presence of X1 or a happy word or to press "l" in the presence of X2 or an unhappy word over the course of Blocks 4 and 5, consisting of 20 trials and 40 trials respectively.

In the following block, the location of the concept labels were reversed (i.e., the label SAM appeared on the top-left side of the screen and the label JIM appeared on the top-right side of the screen). Participants now had to press "a" in the presence of X2 or press "l" in the presence of X1 for a correct response to be recorded (over 20 trials). Blocks 6 and 7 comprised 20 trials and 40 trials, respectively, and involved presenting the HAPPY and UNHAPPY attribute labels with the labels SAM and JIM so that pressing "a" in the presence of X2 or happy words, or pressing "l" in the presence of X1 or unhappy words, were deemed correct responses (see Fig. 2 for an illustration of the block types across the practice IAT). These blocks will be referred to as B1, B2, B3, B4, B5, B6, and B7 in the Results section.

Phase 1a—More/Less Training and Testing

The goal of Phase 1a was to establish Cue 1 and Cue 2 as functionally equivalent to the phrases *more than* and *less than*, respectively, for all participants. The phase involved a 60-trial training block(s) followed by a 12-trial test block. The following instructions were then presented on screen.
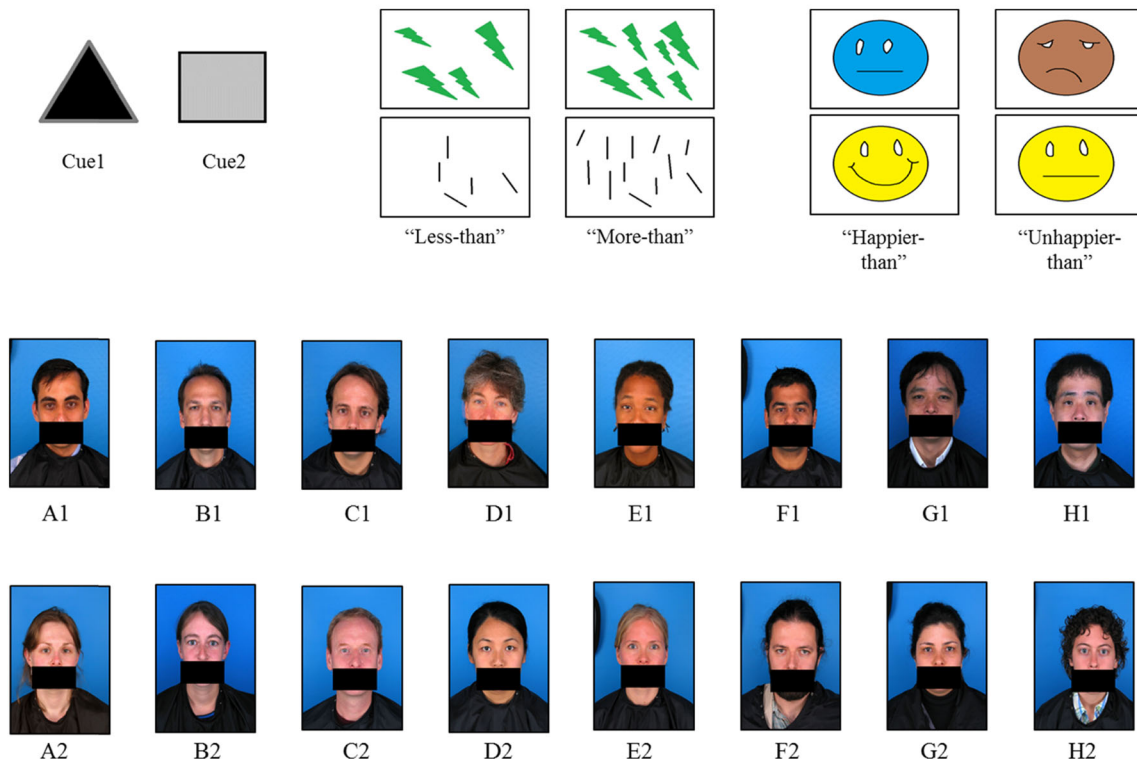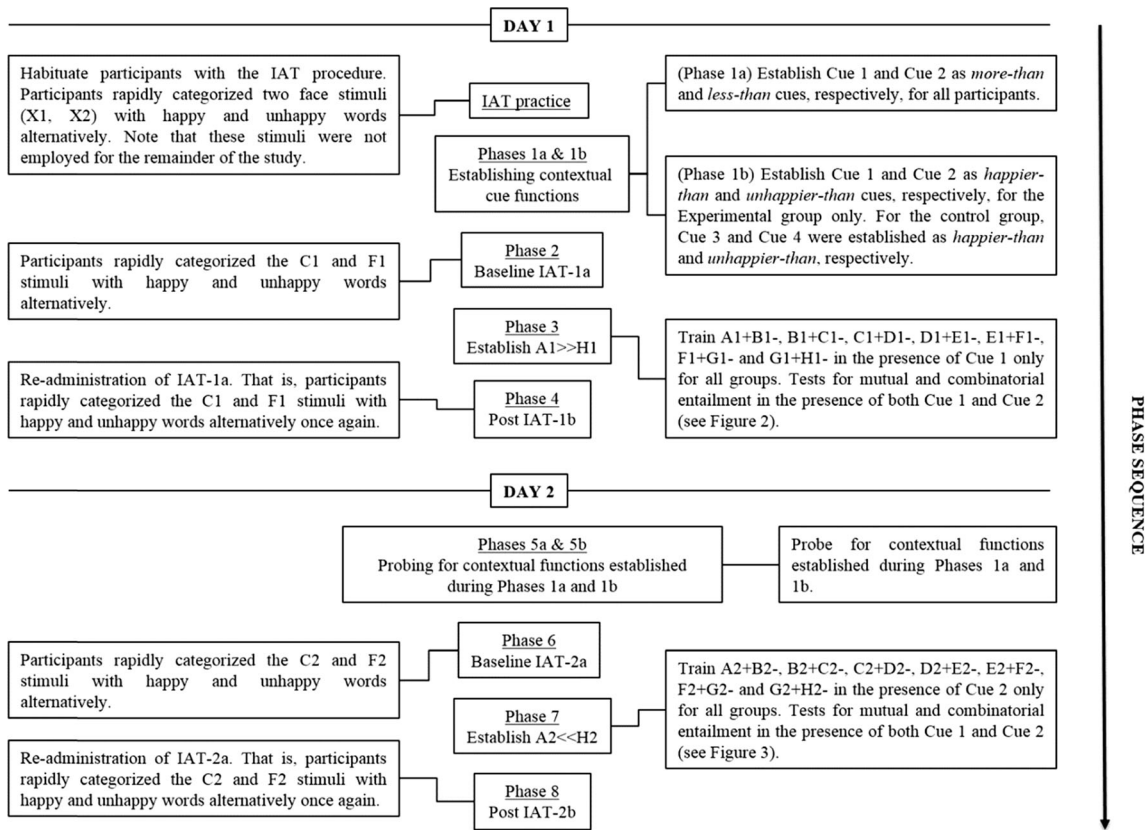
> Welcome! In this phase, you will briefly see a TRIANGLE or a SQUARE come up on the screen, followed by TWO items near the bottom of the screen. One of the items near the bottom will be the "correct" choice. To select the item on the left, please press "a". To select the item on the right, please press "l." You will make some errors at first and that is okay—you will be provided with feedback as you progress. If you have any questions, please ask the experimenter. Otherwise, you may press any key to begin . . .

After pressing any key on the keyboard, the trial initiated with a blank, white screen for 300 ms. This was followed by the appearance of either Cue 1 or Cue 2 against a white background on the top half of the computer screen for 2,000 ms. Immediately after, the cue was removed while stimuli from the *more than* (M1–M4) and *less than* (L1–L4) categories appeared on the left and right sides of the bottom half of the screen and remained there until participants produced the appropriate response—that is, pressed the "a" or "l" key to select the stimulus on the left or right sides of the screen, respectively. The positions of the *more-than* and *less-than* exemplars were counterbalanced across trials. Responses deemed correct were followed by the word "Correct!" presented in a green font in the center of the screen for 1,500 ms, otherwise the word "Wrong" was presented in a red font for 1,500 ms. The mastery criterion for the 60-trial training block was 20 consecutive correct responses. If criterion was not met, participants were re-exposed to the training block until it was. Once criterion was met, participants progressed on to the test trials.

Test trials were similar to the preceding training trials, with the exception that all responses were followed by a blank screen for 1,500 ms before proceeding to the next trial. The completion criterion for the test Phase was set at 11 correct responses out of a 12-trial block. It should be noted that the exemplars employed during the test block incorporated novel comparison-pairs (e.g., M5-L5, M6-L6) that had not been presented during training. Not meeting criterion led to a re-exposure to the training block for a maximum of three times, which was not required as all participants met criterion.

Phase 1b—Happier/Unhappier Training and Testing

The goal of Phase 1a was to establish Cue 1 and Cue 2 as functionally equivalent to the phrases *happier than* and

DAY 1

Habituate participants with the IAT procedure. Participants rapidly categorized two face stimuli (X1, X2) with happy and unhappy words alternately. Note that these stimuli were not employed for the remainder of the study.	IAT practice	(Phase 1a) Establish Cue 1 and Cue 2 as *more-than* and *less-than* cues, respectively, for all participants.
	Phases 1a & 1b Establishing contextual cue functions	(Phase 1b) Establish Cue 1 and Cue 2 as *happier-than* and *unhappier-than* cues, respectively, for the Experimental group only. For the control group, Cue 3 and Cue 4 were established as *happier-than* and *unhappier-than*, respectively.
Participants rapidly categorized the C1 and F1 stimuli with happy and unhappy words alternately.	Phase 2 Baseline IAT-1a	
Re-administration of IAT-1a. That is, participants rapidly categorized the C1 and F1 stimuli with happy and unhappy words alternately once again.	Phase 3 Establish A1>>H1	Train A1+B1-, B1+C1-, C1+D1-, D1+E1-, E1+F1-, F1+G1- and G1+H1- in the presence of Cue 1 only for all groups. Tests for mutual and combinatorial entailment in the presence of both Cue 1 and Cue 2 (see Figure 2).
	Phase 4 Post IAT-1b	

DAY 2

	Phases 5a & 5b Probing for contextual functions established during Phases 1a and 1b	Probe for contextual functions established during Phases 1a and 1b.
Participants rapidly categorized the C2 and F2 stimuli with happy and unhappy words alternately.	Phase 6 Baseline IAT-2a	
Re-administration of IAT-2a. That is, participants rapidly categorized the C2 and F2 stimuli with happy and unhappy words alternately once again.	Phase 7 Establish A2<<H2	Train A2+B2-, B2+C2-, C2+D2-, D2+E2-, E2+F2-, F2+G2- and G2+H2- in the presence of Cue 2 only for all groups. Tests for mutual and combinatorial entailment in the presence of both Cue 1 and Cue 2 (see Figure 3).
	Phase 8 Post IAT-2b	

PHASE SEQUENCE

Cue1   Cue2   "Less-than"   "More-than"   "Happier-than"   "Unhappier-than"

A1   B1   C1   D1   E1   F1   G1   H1

A2   B2   C2   D2   E2   F2   G2   H2

unhappier than, respectively, for participants in the Experimental group only. The training and testing procedures were similar in all respects to those described for Phase 1a, with the exception that the target stimuli

◄ **Fig. 1 a** Sequence of phases in the present experiment. **b** Stimuli used in the current experiment. Specifically, the two shapes employed as contextual cues throughout the study (top left); examples of "more-than" and "less-than" comparisons during Phase 1a (top middle); examples of "happier-than" and "unhappier-than" comparisons during Phase 1b (top right). The A1-H1 stimuli were all employed during Phase 3, where the C1 and F1 stimuli were contrasted across IATs 1a and 1b (middle). The A2-H2 stimuli were all employed during Phase 7, where the C2 and F2 stimuli were contrasted across IATs 2a and 2b.

consisted of various smiley faces instead of different quantities of various objects (see Fig. 1b for examples). Specifically, each participant selected between comparison pairs displaying "very happy" (V++), "happy" (V+), "unhappy" (V-) and "very unhappy" (V–) smiley faces.

For the Experimental group, selecting the happier face in the presence of Cue 1, or the unhappier face in the presence of Cue

2, was reinforced. The emission of 20 consecutively correct discriminations during training was followed by the test block, during which participants had to emit 11 correct responses in order to progress to the next stage. Participants in the Control group underwent a similar procedure to establish Cue 3 and Cue 4 as *happier than* and *unhappier than*, respectively (see Table 1 for individual performances in Phases 1a and 1b).

Phase 2—IAT-1a

All participants underwent an IAT similar to that employed in the practice phase, except that different labels and stimuli were used for the concept categories. Specifically, the labels JIM and SAM (from the practice IAT) were replaced by the labels JOHN and SEAN, and the face stimuli X1 and X2 were
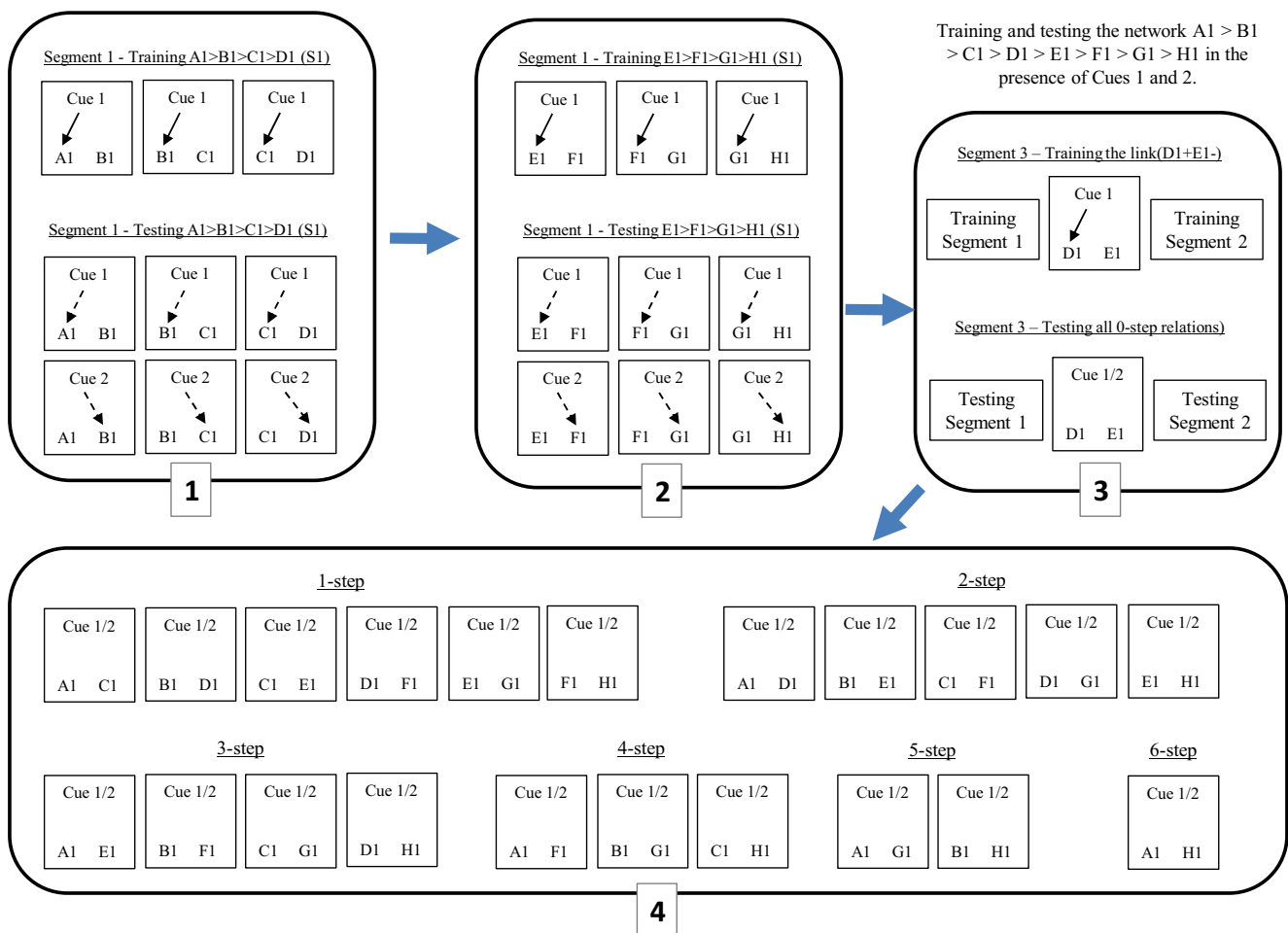

Fig. 2 Segment sequence in Phase 3. In Segment [1], participants were trained the A1+B1-, B1+C1- and C1+D1- discriminations in the presence of Cue 1, where solid arrows indicate the reinforced choice. For example, in the first training trial in Segment 1, selection of A1 was reinforced when A1-B1 appeared in the presence of Cue 1. Following training, participants had to select from the same comparison pairs in the presence of both Cue 1 and Cue2 without any differential feedback where dashed arrows indicate the "correct" choice. In Segment [2], participants were trained the E1+F1-, F1+G1- and G1+H1- discriminations in the presence of Cue 1 only. Testing trials involved the appearance of the three comparison pairs in the presence of both Cue 1 and Cue 2. Segment [3] training involved retraining all participants the conditional discriminations from Segments 1 and 2, as well as the novel D1+E1- discrimination, which "linked" the two A1>B1>C1>D1 and E1>F1>G1>H1 networks. Testing involved all test trials from Segments 1 and 2 as well as the D1-E1 pair. Upon meeting criterion in Segment 3, all participants underwent tests for 1, 2, 3, 4, 5, and 6-step combinatorial entailment in Segment [4].

Table 1 Individual Training and Testing Performances Across Phases 1a and 1b

More Than vs. Less Than (Phase 1a)

ID*	Training trials**	Test trials [1a][+]	Test trials [5a][++]
P1	59	11 [1]	12 [1]
P2	104	12 [1]	12 [1]
P3	62	12 [1]	12 [1]
P4	29	12 [1]	12 [1]
P5	48	12 [1]	12 [1]
P6	38	12 [1]	12 [1]
P7	32	12 [1]	12 [1]
P8	50	12 [1]	12 [1]
P9	64	12 [1]	12 [1]
P10	58	11 [1]	12 [1]

Happier Than vs. Unhappier Than (Phase 1b)

ID	Training trials	Test trials [1b]	Test trials [5b]
P1	19	11 [1]	11 [1]
P2	154	12 [3]	12 [1]
P3	65	12 [1]	12 [1]
P4	142	12 [1]	12 [1]
P5	37	12 [1]	12 [1]
P6	61	12 [1]	12 [1]
P7	119	12 [1]	12 [1]
P8	150	12 [2]	12 [1]
P9	64	12 [2]	12 [1]
P10	58	11 [1]	12 [1]

*ID indicates the individual participant involved—underlined IDs (e.g., P7] indicate participants from the Control group.

** Number of training trials participants underwent prior to reaching the mastery criterion

[+] Number of correct responses produced over 12-trial test blocks in Phase 1a with total number of attempted test blocks presented in brackets. For example, P1 emitted 12 correct responses within one test block, hence 12 [1].

[++] Number of correct responses produced over the 12-trial test block in phase 5a with total number of attempted test blocks presented in brackets.

replaced by the novel C1 and F1 face stimuli. All other parameters remained the same. Note that the sequence of pairings within an IAT (e.g., C1 w/happy words and F1 w/unhappy words followed by F1 w/happy words and C1 w/unhappy words) was counterbalanced between participants. The current IAT, which will hereafter be referred to as IAT-1a, served as a baseline measure for determining how fluently C1 was paired with happy/unhappy, words relative to F1. Completion of IAT-1a was followed by a 5 minute rest interval.

Phase 3—Establish A1>>H1

The goal of the present phase was to establish an eight-member network, A1>B1>C1>D1>E1>F1>G1>H1, where ">" refers to a relation of *happier than* (see Fig. 2 for details

regarding training and testing sequence). Specifically, the phase involved reinforcing the conditional discriminations A1+B1-, B1+C1-, C1+D1-, D1+E1-, E1+F1-, F1+G1-, and G1+H1- in the presence of the *happier-than* cue (Cue 1) only. Alternatively, non-reinforced tests for entailment/transitivity were carried out in the presence of both Cue 1 and Cue 2.

Recall that for participants in the Control group, Cue 1 and Cue 2 were only established as *more-than* and *less-than* cues, respectively. This allowed us to determine whether training with comparative contextual cues with *specified* histories (i.e., Cue 1→*happier than*) yields a more robust transformation of valence functions than training with comparative contexts with more "generic" histories (e.g., Cue 1→*more than*; I will address this issue in greater detail in the Discussion).

Phase 3 consisted of four segments (see Fig. 3). Each segment commenced with the instructions similar to those employed in Phase 1a. That is, participants were informed that they would be presented with a cue, followed by a comparison pair, from which they had to select a comparison stimulus by pressing the appropriate key ("a" for the stimulus on the left, "l" for the stimulus on the right) in order to continue (see Table 2 for details regarding Phase 3).

Segment 1 Participants were shown comparison pairs A1-B1, B1-C1 and C1-D1 in the presence of Cue 1 only. Participants were trained (i.e., given corrective feedback) with respect to the A1+B1-, B1+C1-, and C1+D1- discriminations in quasi-random order in the presence of Cue 1 until 20 consecutively correct responses were recorded. On reaching this criterion, participants progressed to the test block where they responded to the same three comparison pair types in the presence of Cue 1 *or* Cue 2 without any corrective feedback (recall that no discriminations were trained in the presence of Cue 2). Participants had to select the "correct" discrimination 11 times out of a 12-trial block or be re-exposed to the training trials. Completion of the segment was followed by a message instructing participants that they could "take a break" if they chose, otherwise they could continue to the next segment by pressing any key on the keyboard.

Segment 2 Participants were next trained in the E1+F1-, F1+G1- and G1+H1- discriminations in quasi-random order in the presence of Cue 1 until 20 consecutively correct responses were produced. This was followed by test trials involving presentations of the E1-F1, F1-G1, and G1-H1 pairs with Cue 1 or Cue 2 in the absence of any feedback in a 12-trial block (see Table 2). The criterion for completion was set at 11 correct responses, the production of which led to the onset of another verbal prompt for participants to "take a break."

Segment 3 Participants were exposed to re-training of the A1+B1-, B1+C1-, C1+D1-, E1+F1-, F1+G1-, and G1+H1- discriminations, along with the novel D1+E1- discrimination,
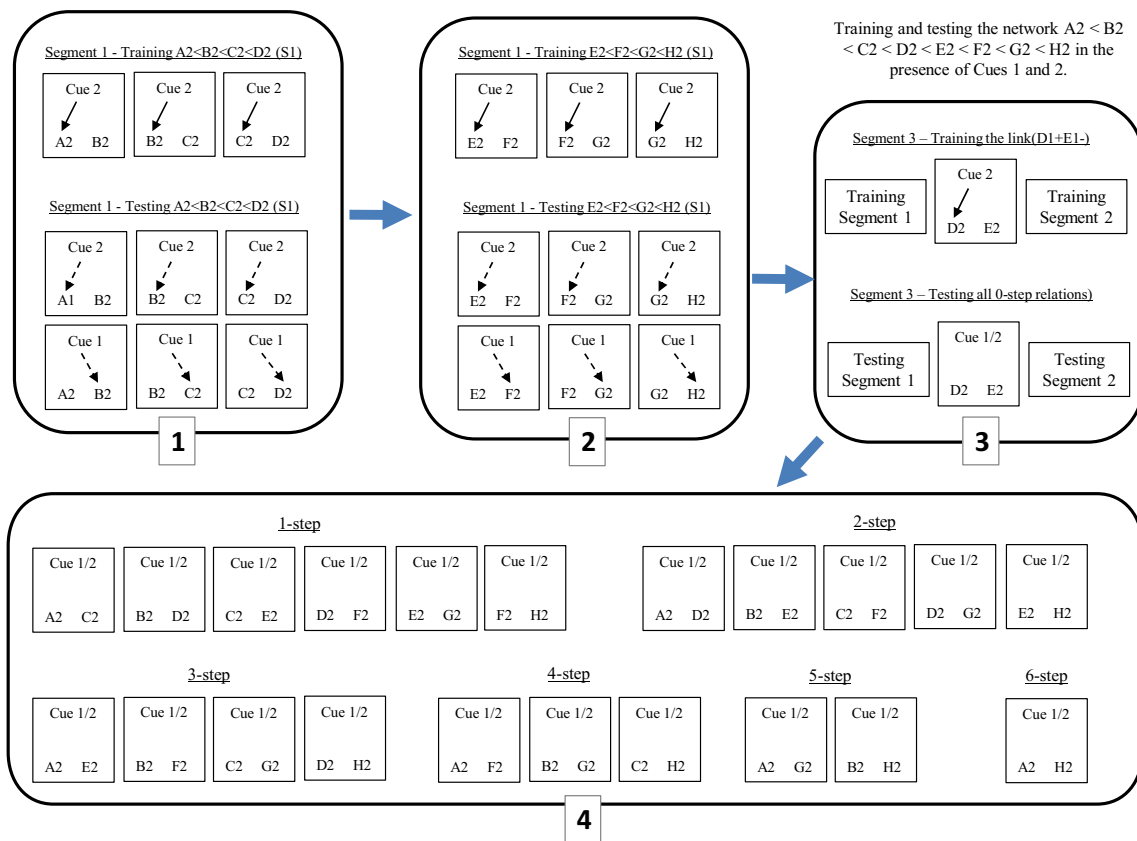
Fig. 3 Segment sequence in Phase 7. Note that the sequence matches exactly that from Phase 3 with the key differences being (a) the comparison pairs employed (i.e., A2-B2-C2-D2-E2-F2-G2-H2) and (b) that only Cue 2 (i.e., *unhappier than*) was used during training trials. Cues 1 and 2 were employed during test trials.

in the presence of Cue 1 only in quasi-random order. Note that training the D1+E1- discriminations "linked" the A1>B1>C1>D1 and E1>F1>G1>H1 networks (i.e., A1>B1>C1>D1>E1>F1>G1>H1). The completion criterion during the training block was the emission of five consecutively correct responses per comparison pair for a total of 40 conditional discriminations. All participants met this criterion. Meeting criterion was followed by the following message[3]:

> During this phase, you will be tested for ALL the relations you have learned so far. You will NOT be provided with feedback at any point during this phase, so please take your time and try to respond as accurately as you can. If you make over a certain number of errors, then you will have to go through the training stage again. Good luck! Press any key to begin . . .

Participants were exposed to all comparison pairs; A1-B1, B1-C1, C1-D1, D1-E1, E1-F1, F1-G1 and G1-H1, in the presence of either Cue 1 or Cue 2 across a 70-trial test block (i.e., with 10 exposures per comparison pair) in quasi-random

order. The completion criterion was 8 correct (out of 10) responses per comparison pair. Failure to meet criterion would lead to re-exposure to training and testing trials for a maximum of three times. All participants eventually met criterion, leading to the final testing segment without any further prompts.

Segment 4 Participants were exposed to a series of comparison pairs in which the stimuli were of various nodal distances, or 'steps', from each other. Specifically, one step (A1-C1, B1-D1, C1-E1, D1-F1, F1-H1), two-step (A1-D1, B1-E1, C1-F1, D1-G1, E1-H1), three-step (A1-E1, B1-F1, C1-G1, D1-H1), four-step (A1-F1, B1-G1, C1-H1), five-step (A1-G1, B1-H1), and six-step (A1-H1) comparison pairs not encountered previously were presented. The pairs were presented in quasi-random order. Each comparison pair appeared four times each in the presence of both Cue 1 and Cue 2, resulting in a 80-trial block. The completion criterion was three correct discriminations (out of four) for each trial type. It was agreed that if participants failed to reach criterion once, they would be reexposed to Segment 4. If they failed a second time, they would be reexposed to Segment 3. If they failed a third time, they would be eliminated from the experiment. This was not necessary for any of the participants. Successful completion was followed by an optional rest interval.

---

[3] Participants P5 and P6 were not provided with this message; they progressed directly to the test trials.

Table 2 Response Accuracies for Phase 3, Segments 1, 2, 3

	Training segment 1[+]			Testing segment 1[++]			
ID*	A1+B1-**	B1+C1-	D1+C1-	B1+A1-	C1+B1-	C1+D1-	
P1	14[16]	13[16]	14[16]	8[8]	8[8]	8[8]	
P2	20[32]	20[32]	18[32]	8[8]	8[8]	8[8]	
P3	28[32]	18[32]	19[32]	10[16]	11[16]	14[16]	
P4	19[32]	18[32]	24[32]	14[16]	14[16]	13[16]	
P5	20[32]	24[32]	20[32]	7[8]	6[8]	7[8]	
P6	18[32]	21[32]	25[32]	10[16]	12[16]	10[16]	
P7	20[32]	13[32]	12[32]	8[8]	8[8]	8[8]	
P8	14[32]	13[32]	20[32]	12[16]	14[16]	13[16]	
P9	17[32]	18[32]	21[32]	8[8]	8[8]	8[8]	
P10	12[16]	14[16]	12[16]	7[8]	8[8]	7[8]	
	Testing segment 1			Training segment 2			
ID	E1+F1-	F1+G1-	G1+H1-	F1+E1-	G1+F1-	H1+G1-	
P1	14[32]	19[32]	21[32]	8[8]	8[8]	8[8]	
P2	12[32]	20[32]	14[32]	9[16]	14[16]	15[16]	
P3	13[16]	14[16]	12[16]	8[8]	8[8]	8[8]	
P4	14[32]	14[32]	20[32]	8[8]	7[8]	7[8]	
P5	10[16]	12[16]	10[16]	7[8]	8[8]	6[8]	
P6	12[16]	12[16]	11[16]	6[8]	7[8]	7[8]	
P7	14[32]	21[32]	12[32]	8[8]	8[8]	8[8]	
P8	12[16]	14[16]	12[16]	8[8]	8[8]	8[8]	
P9	12[32]	20[32]	18[32]	8[8]	7[8]	8[8]	
P10	12[32]	14[32]	20[32]	8[8]	8[8]	8[8]	
Training segment 3							
ID	A1+B1-*	B1+C1-	D1+C1-	D1+E1-****	E1+F1-	F1+G1-	G1+H1-
P1	18[32]	24[32]	19[32]	46[64]	22[32]	23[32]	22[32]
P2	20[32]	18[32]	26[32]	29[64]	26[32]	24[32]	18[32]
P3	14[16]	15[16]	15[16]	24[32]	15[16]	10[16]	14[16]
P4	15[16]	12[16]	14[16]	20[32]	15[16]	10[16]	10[16]
P5	18[32]	19[32]	18[32]	51[64]	29[32]	26[32]	28[32]
P6	12[16]	10[16]	12[16]	21[32]	14[16]	14[16]	13[16]
P7	25[32]	26[32]	19[32]	29[64]	30[32]	27[32]	24[32]
P8	14[16]	13[16]	15[16]	21[32]	15[16]	12[16]	9[16]
P9	18[32]	24[32]	19[32]	46[64]	22[32]	23[32]	22[32]
P10	20[32]	18[32]	26[32]	29[64]	26[32]	24[32]	18[32]
Testing segment 3							
ID	B1+A1-	C1+B1-	D1+C1-	E1+D1-	F1+E1-	G1+F1-	H1+G1-
P1	9[10]	10[10]	10[10]	10[10]	10[10]	10[10]	10[10]
P2	12[20]	18[20]	18[20]	9[20]	17[20]	16[20]	17[20]
P3	10[10]	8[10]	8[10]	9[10]	10[10]	10[10]	9[10]
P4	14[20]	18[20]	15[20]	10[20]	17[20]	18[20]	9[20]
P5	8[10]	8[10]	9[10]	10[10]	10[10]	9[10]	10[10]
P6	11[20]	14[20]	11[20]	17[20]	14[20]	19[20]	18[20]
P7	16[20]	12[20]	17[20]	9[20]	17[20]	11[20]	16[20]
P8	16[20]	20[20]	18[20]	15[20]	13[20]	20[20]	18[20]
P9	12[20]	18[20]	17[20]	16[20]	17[20]	16[20]	17[20]
P10	10[10]	10[10]	9[10]	8[10]	9[10]	8[10]	10[10]

* Indicates individual participants, where underlined IDs (e.g., P8) indicate participants in the control group.

** Comparison pairs presented during the training and testing segments, where (+) indicates the correct response. For example, A1+B1- indicates that selecting A1 was the correct response when A1-B1 appeared in the presence of Cue 1.

*** Number of correct responses produced with the total number of responses produced, for a particular comparison pair, provided in brackets.

**** The D1-E1 pair was presented twice the number of times the other comparison pairs were presented during both training and testing Segment 3.

[+] Comparison pairs employed during the training segment (in the presence of Cue 1 only).

[++] Comparison pairs employed during the testing segment (in the presence of Cue 1 and Cue 2).

Phase 4—IAT-1b

Participants were exposed to a second IAT (IAT-1b) identical to the one employed in Phase 2. The goal was to note any changes in performance compared to baseline scores on IAT-1a. Completion of the IAT signalled the end of the first day (condition) of the procedure. Participants were asked to return within 48 hours to complete the second stage of the study.

Day/Condition 2

Phase 5a—More/Less Probes

Participants were probed for the *more*-than and *less-than* functions established during Phase 1a. Specifically, all participants underwent 12 test trials (from Phase 1a) in which they had to respond "correctly" (select the greater quantity of items in the presence of Cue 1, and select the fewer quantity of items in the presence of Cue 2) at least 11 times. All participants met criterion within the first block.

Phase 5b—Happier/Unhappier Probes

Participants were probed for the *happier-than* and *unhappier-than* functions established during Phase 1b. Specifically, participants in the Experimental group were exposed to 12 test trials (from Phase 1b), from which they had to respond "correctly" (i.e., select the happier face stimulus in the presence of Cue 1, and the unhappier face stimulus in the presence of Cue 2). Participants from the Control group were probed for the *happier-than* and *unhappier-than* functions established for Cues 3 and 4. All participants met criterion of 11 correct responses within the first 12 trials. Completion of Phases 5a and 5b prompted a rest interval.

Phase 6—IAT-2a

Participants were exposed to a second, baseline IAT (IAT-2a), where the JOHN and SEAN labels (from IATs 1a and 1b) were replaced with STAN and JAMES. Concurrently, the C1 and F1 stimuli were replaced with C2 and F2. The goal of the present Phase was to determine a baseline "preference" of one stimulus in relation to the other. The order of C2-F2 pairings with happy-unhappy (or "unhappy-happy") words were counterbalanced across participants. Completion was followed by a 5-minute interval.

Phase 7—Establish A2<<H2

Phase 7 resembled Phase 3 in all respects except for two critical differences. First, a new set of comparison pairs were employed for conditional discrimination training (i.e., A2-

B2, B2-C2, C2-D2, D2-E2, E2-F2, F2-G2, G2-H2; see Fig. 1b). Second, all conditional discriminations were trained in the presence of Cue 2 (trained previously as *unhappier than*) with tests for entailment involving the appearance of comparison pairs in the presence of both Cues 1 and 2. Four subsets of training and testing segments were involved, similar to Phase 3 (see Table 3). Failing to acquire mastery over any set of training and testing segments led to re-exposure to the relevant segments. As the details of the procedure have been outlined previously (see Phase 3), and as all participants eventually met criterion, the training and testing segments have not been discussed further.

Phase 8—IAT-2b

Participants completed an IAT (IAT-2b) that was identical to IAT-1b from Phase 6. Completion of the IAT marked the end of the experiment.

Results

Day/Condition 1

IAT Practice Phase

The function of the IAT practice phase was to habituate participants with the IAT procedure; hence no predictions were made and no results have been presented.

Table 3 Response Accuracies for Phase 3, Segment 4 (for Experiments 1 and 2)

ID	1-step*	2-step	3-step	4-step	5-step	6-step
P1	19[60]	21[60]	29[48]	20[36]	18[24]	10[12]
P2	27[40]	29[40]	18[32]	21[24]	15[16]	6[8]
P3	18[20]	18[20]	15[16]	12[12]	6[8]	4[4]
P4	20[40]	25[40]	26[32]	20[24]	10[16]	7[8]
P5	17[20]	18[20]	11[16]	10[12]	6[8]	3[4]
P6	29[40]	22[40]	25[32]	20[24]	13[16]	6[8]
P7	21[40]	28[40]	26[32]	20[24]	10[16]	6[8]
P8	19[40]	21[40]	18[32]	15[24]	7[16]	6[8]
P9	17[20]	16[20]	14[16]	11[12]	7[8]	4[4]
P10	19[20]	18[20]	16[16]	11[12]	8[8]	4[4]

* Categorizes comparison pairs (for tests of combinatorial entailment) based on internodal distance; specifically, the categories are 1-step (A1-C1, B1-D1, C1-E1, D1-F1, F1-H1), 2-step (A1-D1, B1-E1, C1-F1, D1-G1, E1-H1), 3-step (A1-E1, B1-F1, C1-G1, D1-H1), 4-step (A1-F1, B1-G1, C1-H1), 5-step (A1-G1, B1-H1) and 6-step (A1-H1).

** Number of correct responses (outside brackets) out of total responses made (within brackets).

Phase 1a—More/Less Training and Testing

All participants reached training criterion (of producing 20 consecutively correct responses) in fewer than 80 trials. During the testing trials, all participants emitted at least 11 correct responses within the first 12-trial block.

Phase 1b—Happier/Unhappier Training and Testing

All participants met the training criterion of 20 consecutively correct responses in fewer than 80 trials. All participants also met the test criterion of at least 11 correct responses within the first 12-trial block.

Phase 2—IAT-1a

Response latencies from blocks B3, B4, B6, and B7 of IAT-1a were collected and analyzed in accordance with the recommendations of Greenwald et al. (2003), which is outlined briefly. First, no data was excluded for analysis (since all recorded latencies were between 300 and under 10,000 ms by virtue of the enforced trial duration; see Gavin et al. 2008). Second, means of "correct" response latencies for blocks B3, B4, B6, and B7, as well as two pooled standard deviations ($SD1$ for all trials in blocks B3 and B6 and $SD2$ for all trials in blocks B4 and B7) were computed. Next, response errors were adjusted for by replacing the latencies for incorrect responses with the block means of correct responses, increased by an increment of 600 ms. Fourth, adjusted latency values from each of the four blocks was averaged and labelled B3μ, B4μ, B6μ, and B7μ. Fifth, two difference values (B6μ-B3μ) and (B7μ-B4μ), were calculated and divided by $SD1$ and $SD2$, respectively. Finally, these two values were averaged to give a difference (D-600) score that ranges between +2 and -2 and indicates the magnitude of the IAT effect (see Greenwald et al. 2003 for details regarding statistical analysis of IAT data).

In order to help the reader interpret this score, recall that all IATs employed here involved two concept categories (let us label them A and B) and two attribute categories (HAPPY and UNHAPPY). A positive D score would indicate that concept A was more quickly and accurately (i.e., fluently) paired with the HAPPY attribute and concept B was more fluently paired with the UNHAPPY attribute. Conversely, a negative D score would indicate the opposite pattern.

Given that IAT-1a was employed as a baseline measure, no formal predictions were made. The performance of participants P1, P2, P3, P4, P5, P8, and P9 produced positive D scores, indicating that C1 was paired more readily with happy words, and F1 with unhappy words. Conversely, P6, P7, and P10 produced negative D scores, indicating that F1 was paired

more readily with happy words, and C1 with unhappy words (see Table 6).

Phase 3—Establish A1>>H1

Segment 1 Participants were trained in the A1+B1-, B1+C1-, and C1+D1- discriminations in the presence of Cue 1, and tested in the presence of Cues 1 and 2. During training, P1 and P10 reached criterion within 48 trials; all other participants reached criterion within 96 trials. In the test segment, P1, P2, P5, P7, P9, and P10 reached criterion within 24 trials. P3, P4, and P8 reached criterion within 48 trials (see Table 2 for details of Segments 1–3 in Phase 3).

Segment 2 Participants were trained in the E1+F1-, F1+G1-, and G1+H1- discriminations in the presence of Cue 1 only, and tested in the presence of Cues 1 and 2. During training, P3, P5, P6, and P8 reached criterion within 48 trials; P1, P2, P4, P7, P9, and P10 reached criterion within 96 trials. During testing, all participants met criterion within 24 trials bar P2, who met criterion within 48 trials.
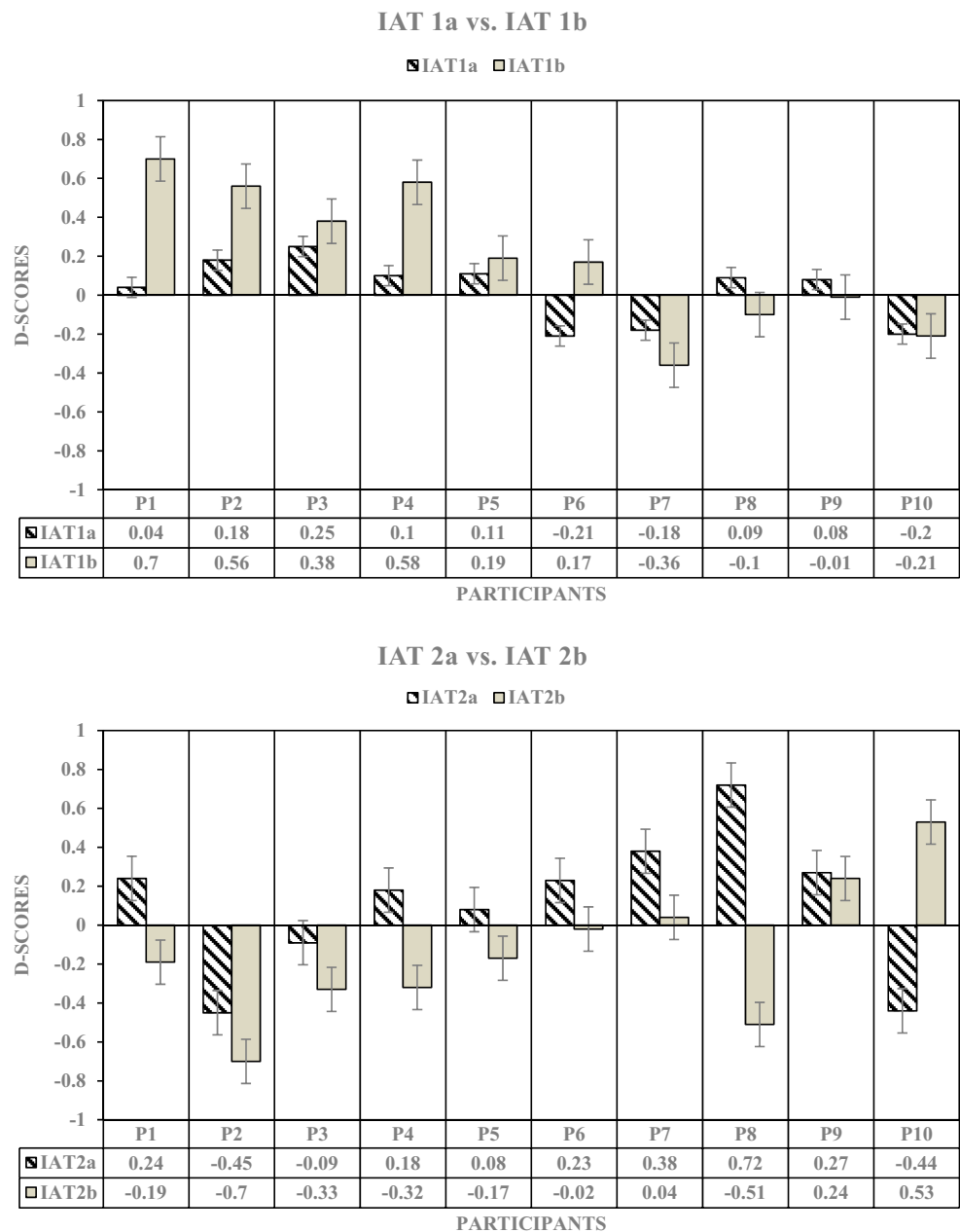
Segment 3 Participants were trained and tested in all discriminations from Segments 1 and 2, along with the novel D1+E1- discrimination. P3, P4, P6, and P8 reached criterion within 128 trials; P1, P2, P5, P7, P9, and P10 reached criterion within 256 trials. During test trials, P1, P3, P5, and P10 reached criterion within 70 trials; P2, P4, P6, P7, P8, and P9 reached criterion within 140 trials.

Segment 4 Participants had to select from one-step, two-step, three-step, four-step, five-step and six-step comparison pairs in the presence of Cue 1 or Cue 2 and in the absence of corrective feedback. All participants eventually reached criterion of producing at least 3 consecutively correct discriminations per comparison pair (see Table 3 for details).

Phase 4—IAT-1b

IAT-1b was a re-administration of IAT-1a. Upon visual comparison of the performances in IAT-1b in relation to IAT-1a, one can immediately observe that all participants in the Experimental group yielded a positive d score in IAT-1b. In contrast, all participants in the Control group yielded a negative d score (see Fig. 4). Critically, the d scores of all six participants in the Experimental group shifted in the predicted direction (see Table 6 and Fig. 4). A Wilcoxon signed-ranks test indicated that, for the Experimental group, IAT-1b ($Mdn=$ 0.11) yielded significantly more positive d scores than IAT-1a ($Mdn=0.47$), $W=0$, $p <0.05$ where $W_{crit}=2$. The z value has not been reported as the size of the sample ($n=6$) was not large enough to form a normal distribution. These performances indicate that the "happiness" (valence) functions of C1 was

Fig. 4 The d scores for IAT-1a vs. IAT-1b (top) and IAT-2a vs. IAT-2b (bottom), where the x-axis indicates the individual participants and the y-axis indicates difference scores. Baseline and post-assessment *d* scores are illustrated by patterned and gray bars, respectively. P1–P6 were members of the Experimental group; P7–P10 were members of the Control group.

IAT 1a vs. IAT 1b

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
IAT1a	0.04	0.18	0.25	0.1	0.11	-0.21	-0.18	0.09	0.08	-0.2
IAT1b	0.7	0.56	0.38	0.58	0.19	0.17	-0.36	-0.1	-0.01	-0.21

PARTICIPANTS

IAT 2a vs. IAT 2b

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
IAT2a	0.24	-0.45	-0.09	0.18	0.08	0.23	0.38	0.72	0.27	-0.44
IAT2b	-0.19	-0.7	-0.33	-0.32	-0.17	-0.02	0.04	-0.51	0.24	0.53

PARTICIPANTS

greater relative to F1 following establishment of the A1>B1>C1>D1>E1>F1>G1>H1 network for all participants in the Experimental group and none in the Control group.

Day/Condition 2

Phase 5a—More/Less Probes

Participants were probed for *more-than* and *less-than* functions established during Phase 1a. All participants met criterion (11 correct out of 12 responses) within a single test block.

Phase 5b—Happier/Unhappier Probes

Participants were probed for *happier-than* and *unhappier-than* properties (established in Phase 1b). All participants met criterion (11 correct out of 12 responses) within a single test block.

Phase 6—IAT-2a

In IAT-2a, participants paired stimuli C2 (STAN) and F2 (JAMES) with happy and unhappy words, alternatively. As IAT-2a was a baseline measure, no formal predictions were

made. Participants P1, P4, P5, P6, P7, P8, and P9 produced positive *d* scores, indicating C2 was paired more readily with happy words, and F2 with unhappy words. Remaining participants P2, P3, and P10 yielded negative *d* scores, indicating C2 was paired more readily with unhappy words, and F2 with happy words.

Phase 7—Establish A2<<H2

Segment 1 Participants were trained and tested in the A2+B2-, B2+C2-, and C2+D2- discriminations in the presence of Cues 1 and 2. During training trials, P2, P4, P7, P8, and P10 met criterion within 48 trials; P1, P3 and P9 met criterion within 96 trials. During test trials, P1, P2, P5, P6, P7, P8, and P10 met criterion within 24 trials; P3, P4, and P9 met criterion within 48 trials (see Table 4 for details of Segments 1–3 in Phase 7).

Segment 2 Participants were trained in the E2+F2-, F2+G2-, and G2+H2- discriminations in the presence of Cue 2 only. P1 met training criterion within 48 trials; P2, P3, and P4 met criterion within 96 trials. During test trials in the presence of Cues 1 and 2, P1, P2, and P3 met criterion within two 12-trial test blocks; P4 met test criterion within four test blocks.

Segment 3 Participants were trained and tested with respect to all discriminations from Segments 1 and 2, along with the novel D2+E2- discrimination. During training trials, P2, P9, and P10 met criterion within 128 trials; P1, P4, P6, P7, and P8 met criterion within 256 trials; P3 and P5 met criterion within 384 trials. During the test block, P1, P2, P5, P9, and P10 met criterion within 70 trials; P3, P4, P6, P7, and P8 met criterion within 140 trials.

Segment 4 All participants reached criterion for combinatorial entailment within three iterations of the test trial blocks. That is, they produced at least 3 correct discriminations per comparison pair (see Table 5).

Phase 8—IAT-2b

In IAT-2b, the performances of all participants in the Experimental group were as predicted. That is, P1-P6 produced *d* scores that were relatively more negative than at baseline (see Table 6). That is, participants paired F2 with happy words, and C2 with unhappy words, more fluently then at baseline. In the Control group, P7, P8, and P9 produced similar performances (although the change in performance for P9 was negligible). A Wilcoxon signed-ranks test indicated that, for the Experimental group, IAT-2b (*Mdn*=0.13) yielded significantly more positive *d*- cores than IAT-2a (*Mdn*=-0.26), *W*=0, *p*<.05 where W_{crit}=2. As before, the *z* value has not been reported as the size of the sample (*n*=6) was too small.

These performances indicate that the "happiness" (valence) functions of C2 was lesser relative to F2 following establishment of the A2<B2<<u>C2</u><D2<E2<<u>F2</u><G2<H2 network for all participants in the Experimental group and three participants in the Control group.

Across both days/conditions of the experiment, the Experimental group produced the predicted performance on 8/8 occasions. The Control group produced the predicted performance on 3/8 occasions. Let us assume that the chance probability (*P*) for a "successful" outcome (i.e., demonstrating the predicted IAT performance) is 0.5 per participant. Let us input the chance probability (*P*=0.5) in the formula for binomial probability, $\binom{n}{k}P^k q^{n-k}$ where *k* is the number of successful outcomes, *P* is the probability of a success per individual outcome (0.5), *n* is the number of outcomes observed (8), and *q* is the probability of failure in one trial (1-*P*). We may then compute the binomial probability of 8/8 successful outcomes for the Experimental group to be *p*<.004. The binomial probability of 3/8 successful outcomes for the Control group would then be *p*<.22. In other words, the probability that the Experimental group produced the observed performance patterns by chance alone is less than four in a thousand.

Discussion

The present experiment involved participants undergoing training and testing sequences designed to establish two geometrical shapes as contextual cues functionally equivalent to the phrases *more than/happier than*, as well as *less than/unhappier than*. Participants were next exposed to a baseline IAT where they had to pair stimuli C1 and F1 with happy and unhappy words, respectively, across two trial blocks and then reverse this categorization pattern (i.e., pair C1 with unhappy and F1 with happy words) in two subsequent trial blocks. Participants were then trained along the A1+B1-, B1+C1-, C1+D1-, D1+E1-, E1+F1-, F1+G1-, and G1+H1- discriminations in the presence of Cue 1 only, followed by tests for mutual and combinatorial entailment in the presence of Cue 1 and Cue 2. The critical between-group manipulation was the training history associated with Cues 1 and 2, in which only the Experimental group received specific training to establish Cue 1 and Cue 2 as *happier than* and *unhappier than*, respectively. Following relational training and testing, the previous IAT was re-administered. IAT performances for all participants in the Experimental group demonstrated that C1 was in fact more readily paired with happy words than F1, relative to baseline performances. On a second day, the procedure described above was replicated and the aforementioned effect was reversed. Specifically, participants were trained the A2+B2-, B2+C2-, C2+D2-, D2+E2-, E2+F2-, F2+G2-, and G2+H2- discriminations in the presence of Cue 2 only. Pre vs. post IAT performances revealed once again that

Table 4 Response Accuracies for Phase 7, Segments 1, 2, 3 (for Experiments 1 and 2)

	Training segment 1†			Testing segment 1††		
ID	A2+B2-*	B2+C2-	D2+C2-	B2+A2-	C2+B2-	C2+D2-
P1	20[32]**	19[32]	24[32]	8[8]	8[8]	8[8]
P2	12[16]	13[16]	14[16]	7[8]	7[8]	7[8]
P3	18[32]	12[32]	14[32]	10[16]	9[16]	11[16]
P4	13[16]	12[16]	9[16]	10[16]	14[16]	16[16]
P5	21[32]	18[32]	20[32]	8[8]	7[8]	8[8]
P6	19[32]	25[32]	21[32]	8[8]	8[8]	7[8]
P7	12[16]	13[16]	14[16]	8[8]	8[8]	8[8]
P8	13[16]	9[16]	14[16]	8[8]	7[8]	8[8]
P9	22[32]	14[32]	18[32]	14[16]	12[16]	13[16]
P10	12[16]	14[16]	11[16]	8[8]	8[8]	8[8]
	Training segment 2			Testing segment 1		
ID	E2+F2-	F2+G2-	G2+H2-	F2+E2-	G2+F2-	H2+G2-
P1	14[16]	13[16]	14[16]	8[8]	8[8]	8[8]
P2	13[32]	14[32]	21[32]	8[8]	8[8]	8[8]
P3	16[32]	20[32]	20[32]	8[8]	8[8]	8[8]
P4	17[32]	14[32]	22[32]	12[16]	9[16]	8[16]
P5	10[16]	12[16]	12[16]	7[8]	8[8]	7[8]
P6	18[32]	19[32]	20[32]	11[16]	12[16]	12[16]
P7	21[32]	14[32]	12[32]	14[24]	21[24]	18[24]
P8	12[16]	13[16]	13[16]	8[8]	8[8]	8[8]
P9	16[32]	18[32]	25[32]	8[8]	8[8]	8[8]
P10	31[48]	28[48]	36[48]	16[24]	14[24]	20[24]

Training segment 3							
ID	A2+B2-*	B2+C2-	D2+C2-	D2+E2- ***	E2+F2-	F2+G2-	G2+H2-
P1	28[32]	27[32]	31[32]	31[64]	26[32]	29[32]	31[32]
P2	16[16]	16[16]	15[16]	21[32]	16[16]	14[16]	16[16]
P3	41[48]	39[48]	40[48]	85[96]	42[48]	39[48]	44[48]
P4	26[32]	27[32]	29[32]	41[64]	24[32]	25[32]	30[32]
P5	34[48]	44[48]	42[48]	84[96]	29[48] 31[48]	38[48]	
P6	26[32]	29[32]	21[32]	59[64]	24[32]	29[32]	27[32]
P7	27[32]	30[32]	31[32]	29[64]	26[32]	29[32]	30[32]
P8	29[32]	27[32]	29[32]	38[64]	22[32]	28[32]	31[32]
P9	15[16]	12[16]	16[16]	22[32]	15[16]	13[16]	16[16]
P10	13[16]	12[16]	15[16]	19[32]	15[16]	13[16]	14[16]

Testing segment 3							
ID	B2+A2-	C2+B2-	D2+C2-	E2+D2-	F2+E2-	G2+F2-	H2+G2-
P1	8[10]	9[10]	9[10]	9[10]	8[10]	9[10]	8[10]
P2	9[10]	8[10]	8[10]	9[10]	8[10]	9[10]	8[10]
P3	17[20]	15[20]	14[20]	13[20]	18[20]	17[20]	16[20]
P4	14[20]	18[20]	14[20]	12[20]	17[20]	14[20]	19[20]
P5	9[10]	10[10]	9[10]	10[10]	9[10]	10[10]	8[10]
P6	15[20]	12[20]	18[20]	13[20]	18[20]	11[20]	14[20]
P7	16[20]	15[20]	15[20]	13[20]	10[20]	14[20]	18[20]
P8	15[20]	18[20]	14[20]	11[20]	16[20]	18[20]	14[20]
P9	8[10]	9[10]	8[10]	9[10]	8[10]	9[10]	8[10]
P10	9[10]	9[10]	8[10]	10[10]	9[10]	9[10]	8[10]

* Comparison pairs presented during the training and testing segments, where (+) indicates the correct response. For example, A2+B2- indicates that selecting A2 was the correct response when A2-B2 appeared in the presence of Cue 2.

** Number of correct responses produced with the total number of responses produced, for a particular comparison pair, provided in brackets.

*** The D2-E2 pair was presented twice the number of times the other comparison pairs were presented during both training and testing Segment 3

† Comparison pairs employed during the training segment (in the presence of Cue 2 only).

†† Comparison pairs employed during the testing segment (in the presence of Cue 1 and Cue 2)

Table 5 Response Accuracies for Phase 7, Segment 4 (for Experiments 1 and 2)

ID	1-step*	2-step	3-step	4-step	5-step	6-step
P1	51[60]	41[60]	39[48]	22[36]	11[24]	7[12]
P2	38[40]	35[40]	29[32]	18[24]	10[16]	5[8]
P3	16[20]	16[20]	10[16]	9[12]	6[8]	3[4]
P4	51[60]	50[60]	42[48]	31[36]	18[24]	4[12]
P5	19[20]	14[20]	10[16]	10[12]	6[8]	3[4]
P6	18[20]	14[20]	9[16]	8[12]	7[8]	4[4]
P7	17[20]	18[20]	14[16]	9[12]	7[8]	3[4]
P8	28[40]	31[40]	21[32]	14[24]	10[16]	5[8]
P9	42[60]	37[60]	28[48]	18[36]	15[24]	7[12]
P10	17[20]	16[20]	12[16]	14[24]	10[16]	5[8]

* Categorizes comparison pairs (for tests of combinatorial entailment) based on internodal distance; specifically, the categories are 1-step (A2-C2, B2-D2, C2-E2, D2-F2, F2-H2), 2-step (A2-D2, B2-E2, C2-F2, D2-G2, E2-H2), 3-step (A2-E2, B2-F2, C2-G2, D2-H2), 4-step (A2-F2, B2-G2, C2-H2), 5-step (A2-G2, B2-H2), and 6-step (A2-H2).

** Number of correct responses (outside brackets) out of total responses made (within brackets).

for all members of the Experimental group, C2 was more readily paired with unhappy words than F2, relative to baseline.

The present article provides robust evidence for the *derivation* of valence functions across comparative relations. To appreciate the significance of this point, recall that at no stage during the present experiment were the face stimuli employed in the relational network directly or indirectly paired with appetitive and/or aversive eliciting stimuli. Our findings highlight how instrumental/operant processes such as 'derivation' and the 'transformation of functions' may be critical in understanding the propagation of emotional functions across symbolic networks (see Dymond et al. 2014, for a similar argument).

One may argue at this point that pairing the C and F stimuli with happy and unhappy words during the baseline IATs could have established the valence functions through primarily Pavlovian means. In order to address this issue, let us examine the sequence of IATs in greater detail. Recall that the sequences of the C1-F1 pairings with happy and unhappy words were counter-balanced across all six participants. It follows then that any significant order effects emerging from baseline

Table 6 Corrected Response Latencies and Difference Scores Across IATs 1a/1b and 2a/2b

IAT-1a vs. IAT-1b (Day 1)

ID	B3+B4[1a]*	B6+B7[1a]	d-score [1a]**	B3+B4[1b]**	B6+B7[1b]	d score [1b]
P1	666	674	0.04	1128	2008	0.70
P2	1161	1302	0.18	496	609	0.56
P3	549	962	0.25	931	1202	0.38
P4	626	656	0.10	582	985	0.58
P5	1307	1502	0.11	609	751	0.19
P6	1585	1278	-0.21	911	1125	0.17
P7	1204	1146	-0.18	602	567	-0.36
P8	1194	1161	0.09	710	750	-0.10
P9	1302	701	0.08	905	901	-0.01
P10	755	798	-0.20	900	784	-0.21

IAT-2a vs. IAT-2b (Day 2)

ID	B3+B4[2a]	B6+B7[2a]	d-score [2a]	B3+B4[2b]	B6+B7[2b]	d score [2b]
P1	617	646	0.27	755	701	-0.19
P2	756	649	-0.45	941	634	-0.70
P3	756	736	-0.09	717	600	-0.33
P4	1161	1302	0.18	696	551	-0.32
P5	994	1087	0.08	1310	1190	-0.17
P6	895	1125	0.23	998	968	-0.02
P7	645	740	0.38	666	674	0.04
P8	553	683	0.72	712	600	-0.51
P9	688	737	0.27	843	904	0.24
P10	901	725	-0.44	678	798	0.53

*Response latencies averaged across blocks B3 and B4 for IAT-1a, hence B3+B4[1a]. The subsequent reported latency is the average across blocks B6 and B7 for IAT-1a, hence B6+B7[1a]. That is, the superscript indicates the IAT from which the illustrated values were taken.

** A positive *d* score indicates that C1 (C2 for IATs 2a and 2b) was paired with happy words more fluently then D1 (D2 for IATs 2a and 2b). A negative *d* score indicates the opposite pattern.

IATs should have produced successful outcomes for exactly half the participants, which was clearly not the case.

Another feature of the current study worth noting was the inclusion of partially masked face stimuli for establishing the relational networks. Pilot research indicated that using these particular stimuli served to facilitate the emergence of the predicted relational responding and transformation effects presently observed. Humans have a long history, both evolutionary and personal, which establishes faces as discriminative for the emotional states of con-specifics (Ekman et al. 1990). This is congruent with other findings from derived relations research which has shown, for example, how equivalence class formation may be facilitated by the inclusion of meaningful stimuli as exemplars (Doran and Fields 2012). By obscuring a critical feature for assessing interpersonal emotions (mouth curvature), one could argue that participants' relational responses were "forced" to come under the control exerted by the contextual cues rather than the topographical features of the face stimuli. In other words, given that (a) the critical features of the face stimuli were masked (see Ekman et al. 1990) and (b) the results of the participants were not homogenous across groups, it is unlikely that facial topography was a determining confound, otherwise the majority of participants would have responded to a stimulus type as being happier/unhappier-than another (which was not the case; see Fig. 4). Perhaps future researchers could explore systematically the extent to which the use of face stimuli facilitates the transformation of function effects observed in the current study.

The reader familiar with derived relational responding processes may wonder if it was necessary to train and test *more-than/less-than* functions in conjunction with *happier-than/unhappier-than* functions when simply establishing the latter should suffice for demonstrating the transformation of function effect (Hayes et al. 2001). Although this issue was not a primary concern of the current experiment, ancillary experimentation that ran parallel to this experiment suggested that training both types of contextual control did in fact yield more robust function transformation effects than training *happier-than/unhappier-than* relations alone. Of course, this question is a good one, and still remains to be addressed empirically in dedicated research. At present, it suffices to suggest that extensive training with multiple exemplars which are comparatively related to one another along numerous dimensions (e.g., more vs. less objects *and* more vs. less emotionality) may occasion more robust contextual control than simply providing training along a single dimension (more vs. less emotionality). Future research could address this inquiry in a number of ways. For instance, one might examine what would happen if cues controlling for *more-than* and *less-than* relations were presented with emotional comparison pairs (e.g. happy vs. unhappy faces) without any training; would contextual control generalize from more vs. less "objects" to more vs. less "emotionality" without training? Would such generalization,

if observed, be reversible? At the very least, such questions could highlight the variables that serve to establish strong versus weaker instances of comparative contextual control over evaluative functions.

In conclusion, the current study demonstrated a *derived* transformation of valence functions across comparatively related stimuli in a very large relational network consisting of up to six-step entailed relations (i.e., eight members). It was also novel insofar as it employed IAT performances as both baseline and post-test measures of emotional valence and valence change following relational training. In so doing these findings make a unique contribution to the behavioristic approach to understanding complex relational responding and how the valence of emotional stimuli transform and propagate in accordance with complex relational networks.

References

Amd, M., & Barnes-Holmes, D. (2014). A derived transformation of evaluative preferences using Implicit Association Tests. *The Psychological Record, 64*, 475–485. **Retrieved from** http://download.springer.com/static/pdf/321/art%253A10.1007%252Fs40732-014-0033-z.pdf?auth66=1421915528_f6f535efbbd514b1cb38c48868ef8123&ext=.pdf.

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

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: experimental evidence for a relational operant. *Journal of Applied Behavior Analysis, 40*(1), 45–71. doi:10.1901/jaba.2007.7-06.

De Almeida, J. H., & de Rose, J. C. (2015). Changing the meaningfulness of abstract stimuli by the reorganization of equivalence classes: effects of delayed matching. *The Psychological Record*. doi:10.1007/s40732-015-0120-9.

Delius, J. D., & Siemann, M. (1998). Transitive responding in animals and humans: exaptation rather than adaptation? *Behavioural Processes, 42*(2/3), 107–137. doi:10.1016/S0376-6357(97)00072-7.

Doran, E., & Fields, L. (2012). All stimuli are equal, but some are more equal than others: measuring relational preferences within an equivalence class. *Journal of the Experimental Analysis of Behavior, 98*(3), 243–256. doi:10.1901/jeab.2012.98-243.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*(3), 331–351. **Retrieved from** http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1334471&tool=pmcentrez&rendertype=abstract.

Dougher, M., Perkins, D. R., Greenway, D., Koons, A., & Chiasson, C. (2002). Contextual control of equivalence-based transformation of functions. *Journal of the Experimental Analysis of Behavior, 78*(1), 63–93. doi:10.1901/jeab.2002.78-63.

Dymond, S., Schlund, M. W., Roche, B., & Whelan, R. (2014). The spread of fear: symbolic generalization mediates graded threat-avoidance in specific phobia. *Quarterly Journal of Experimental Psychology, 67*(2), 247–259. doi:10.1080/17470218.2013.800124.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: emotional expression and brain physiology. II. *Journal of Personality and Social Psychology, 58*(2), 342–353. **Retrieved from** http://www.ncbi.nlm.nih.gov/pubmed/2319446.

Gavin, A., Roche, B., & Ruiz, M. R. (2008). Competing contingencies over derived relational responding: a behavioral model of the Implicit Association Test. *The Psychological Record, 58*, 427–441.

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. **Retrieved from** http://www.ncbi.nlm.nih.gov/pubmed/9654756.

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.

Guez, D., & Audley, C. (2013). Transitive or not: a critical appraisal of transitive inference in animals. *Ethology, 119*(9), 703–726. doi:10.1111/eth.12124.

Hayes, S. C., Barnes-holmes, D., & Roche, B. (2001). *Relational frame theory: A Post-Skinnerian account of human language and cognition*. New York, NY: Springer. **Retrieved from** http://books.google.com/books?hl=en&lr=&id=n4RmapzrihAC&pgis=1.

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

Martin, N., & Alsop, B. (2004). Transitive inference and awareness in humans. *Behavioural Processes, 67*(2), 157–165. doi:10.1016/j.beproc.2004.03.017.

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.

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.

Ridgeway, I., Roche, B., Gavin, A., & Ruiz, M. R. (2010). Establishing and eliminating Implicit Association Test effects in the laboratory: Extending the behavior-analytic model of the IAT, 2(2), 133–150.

Roche, B., & Dymond, S. (2008). A transformation of functions in accordance with the nonarbitrary relational properties of sexual stimuli. *The Psychological Record, 58*, 71–90.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Sharpsburg, PA: Psychology Software Incorporated. **Retrieved from** https://scholar.google.com/scholar?hl=en&as_sdt=0,5&cluster=5372168343504638999#0.

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56*(3), 519–555. doi:10.1901/jeab.1991.56-519.

Vasconcelos, M. (2008). Transitive inference in non-human animals: an empirical and theoretical analysis. *Behavioural Processes, 78*(3), 313–334. doi:10.1016/j.beproc.2008.02.017.

Weaver, J. E., Steirn, J. N., & Zentall, T. R. (1997). Transitive inference in pigeons: control for differential value transfer. *Psychonomic Bulletin & Review, 4*(1), 113–117. doi:10.3758/BF03210782.