CrossMark

A derived transformation of emotional functions using self-reports, implicit association tests, and frontal alpha asymmetries

Micah Amd[1] · Bryan Roche[1]

Abstract Research on the derived transformation of stimulus functions (ToF) typically employs single dependent measures for assessing the stimulus functions after derived relations have been established. For the first time, we examined ToF using three dependent measures both prior to and after relational training and testing. Specifically, we employed self-reports, implicit association tests, and frontal alpha asymmetry as pre versus post measures for assessing ToF. First, we trained two abstract shapes as contextual cues for *happier-than* and *unhappier-than* relations, respectively. Next, four conditional discriminations (A+/B–, B+/C–, C+/D–, and D+/E–) were trained in the presence of the *happier-than* cue only, where A, B, C, D, and E were blurred faces. This was followed by tests for contextually controlled transitive inference (TI) in the presence of both the *happier-than* and *unhappier-than* cues. For the participants who demonstrated TI, performance across all three measures following relational training and testing indicated that the "happiness" functions of the A/B stimuli were greater than those of the D/E stimuli. This constitutes the first known demonstration of emotional ToF along explicit, implicit, and neurophysiological measures concurrently.

Keywords Transformation of functions · Frontal alpha asymmetry · Transitive inference · Symbolic relations

Various organisms can infer a relation between a wide array of stimuli given a common, mediating stimulus, and in so doing demonstrate what has been called a "transitive inference" (TI;

✉ Micah Amd
   micah.amd.eab@hotmail.com

[1]   Maynooth University, Maynooth, Ireland

Vasconcelos, 2008). To illustrate TI in a human, imagine we provide a participant with the following statements—"Adam is happier than Bob," "Bob is happier than Carry," "Carry is happier than Dan," and "Dan is happier than Eden." If he or she were then asked to select the "happier" individual from between Bob and Dan, selecting Bob would be deemed a demonstration of TI, since to provide the correct answer required making an inference based on the common node (Carry) that Bob and Dan are both related to (i.e., if Bob $>_{happy}$ Carry and Carry $>_{happy}$ Dan, then Bob $>_{happy}$ Dan). Although many similarities exist between human and nonhuman TI performances (Martin & Alsop, 2004), one feature of TI that may be unique to verbally sophisticated humans is the so-called "transformation of (stimulus) functions," or ToF, effect (Dymond & Rehfeldt, 2000). ToF was first elaborated in an account of symbolic stimulus–stimulus relations known as *relational frame theory* (see Hayes, Barnes-Holmes, & Roche, 2001; but see also Clark, 1969). Simply stated, ToF describes the alteration, or *transformation*, of the psychological functions of a stimulus in accordance with its relationship to (at least) another stimulus. Following from our previous example, deriving Bob as being happier than Dan (Bob $>_{happy}$ Dan) following training in the Bob $>_{happy}$ Carry and Carry $>_{happy}$ Dan pairs may transform the "happiness" functions of the related stimuli so that Bob *becomes* "happier" (more positively valenced) than Dan as a consequence of deriving Bob $>_{happy}$ Dan (Amd & Barnes-Holmes, 2014; Amd & Roche, 2015).

One way of testing for ToF of the type described in our example above could involve requiring a participant to rapidly categorize the Bob and Dan stimuli with happy and unhappy words alternatively, as in the implicit association test format (IAT; Greenwald, McGhee, & Schwartz, 1998). Briefly, the IAT requires individuals to categorize stimuli into pairs across two testing blocks, in which one block involves producing

historically congruent categorizations (e.g., the picture of a flower with the word "pleasant") and a second block involves historically incongruent categorizations (e.g., the picture of a dirty toilet with the word "pleasant"). In light of our previous example, imagine that the Bob and Dan stimuli are presented in the context of an IAT format in which they have to be rapidly categorized with happy and unhappy words alternatively. If we observe that our hypothetical participant is pairing Bob more fluently with happy words, and Dan more fluently with unhappy words, we can assume that Bob is *functioning* as "happier" than Dan (for demonstrations of ToF using IATs, see Amd & Barnes-Holmes, 2014; Amd & Roche, 2015; O'Toole, Barnes-Holmes, & Smyth, 2007).

The IAT effect described above was recently demonstrated in a study by Amd and Barnes-Holmes (2014). In the study, participants were first exposed to different emotional face pairs (happy–unhappy, neutral–happy, and unhappy–neutral) presented in either a yellow or a red background context. If the background was yellow, selecting the relatively happier face was rewarded—alternatively, if the background was red, selecting the relatively unhappier face was rewarded. Tests for generalization of the performance to novel stimulus pairs affirmed that the contextual functions of the red and yellow backgrounds were functioning as *unhappier-than* and *happier-than*, respectively. Participants were then trained on three overlapping discriminations (A+/B–, B+/C–, C+/D–) in the presence of a yellow background only. Next, they underwent tests for contextually controlled TI. This involved presenting comparison pairs that had been employed during training trials (e.g., B–C), but in the presence of the red background (which had not been employed during training) for what was deemed tests for functional *symmetry*. Across such tests, participants typically showed a reversal of their stimulus selections, as we expected given the contextual cue present (e.g., selecting C rather than B in the presence of *unhappier than* in a test trial after being trained to select B rather than C in the presence of *happier than*). Participants were also tested for functional *transitivity*, which involved presenting pairs of stimuli that had never appeared together during training, such as B–D, on both yellow and red backgrounds across separate trials. Participants who chose the "happier" of the pair in the presence of the yellow background and the "unhappier" of the pair in the presence of the red background, as expected, were deemed to have exhibited TI in accordance with the relational series established (i.e., A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D). Subsequent testing for ToF involved exposing participants two IATs for which, in the first IAT, participants had to rapidly categorize the A and D stimuli with happy and unhappy words across alternate blocks. In the second IAT, participants had to categorize the B and C stimuli with happy and unhappy words across alternate blocks. For participants who demonstrated TI, stimulus A was more readily categorized with happy words than was stimulus D in IAT 1, whereas stimulus B was more readily categorized with happy words than was stimulus C in IAT 2. In effect, the valence functions of the stimuli transformed in accordance with the established A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D hierarchy, as we confirmed using a novel IAT format.

One goal of the present study was to extend the findings of Amd and Barnes-Holmes (2014) by incorporating other dependent measures alongside IATs performed both before and after testing for TI. The first measure was a 5-point Likert scale that allowed participants to rate the happiness level of a stimulus from *happy* to *neutral* to *unhappy*. Including a self-report measure allowed us to directly assess whether participants' appraisals of stimulus "happiness" were affected by the experimental procedure. A second dependent measure employed was an electrophysiological measure known as *frontal alpha asymmetry* (FAA). FAA refers to the power difference along an "alpha" bandwidth between the left and right anterior brain regions (Bazanova & Vernon, 2014; Coan & Allen, 2004; Sutton & Davidson, 1997). FAAs have been used to discriminate positive versus negative affect in humans from as early as 2 years of age (Fox & Davidson, 1986). Specifically, increased levels of alpha activation over the left hemisphere indicate positive/approach-related affect, whereas increased alpha activation over the right hemisphere indicates negative/withdrawal-related affect (Briesemeister, Tamm, Heine, & Jacobs, 2013).

To date, no published study has examined whether FAAs are sensitive to derived emotional effects established across comparatively related stimuli. The first study to use electroencephalography (EEG) measures to assess emotional ToF effects within the context of derived stimulus relations employed event-related potentials (ERPs) as a dependent measure (Amd, Barnes-Holmes, & Ivanoff, 2013). Specifically, Amd et al. trained participants to establish three, three-member equivalence classes (A1–B1–C1, A2–B2–C2, A3–B3–C3), where the A1, A2, and A3 stimuli had been paired with emotionally positive, neutral, and negative pictures, respectively. Subsequent presentations of the A1, A2, and A3 stimuli elicited EEG effects indicative of differential emotional valences, as would be expected for symbolic stimuli directly paired with emotion-eliciting images (cf. Hinojosa, Carretié, Valcárcel, Méndez-Bértolo, & Pozo, 2009). The novel finding, however, was that the ERPs elicited by the C1, C2, and C3 stimuli (which had never been presented with the A stimuli) appeared topographically similar to those elicited by the A1, A2, and A3 stimuli, respectively. The similarities in ERP modulation across the A and C stimuli led Amd et al. to argue that the valence functions of the A stimuli had transferred to the C stimuli.

Although the experiment by Amd et al. (2013) was the first published attempt to utilize EEG in assessing emotional function transfer, some methodological limitations warrant mention. First, note that Amd et al. made no baseline EEG measurements of the A and C stimuli, making it difficult to ascertain whether

the observed EEG effects were directly caused by the experimental procedure. Given the large number of trials required to acquire ERPs (Light et al., 2010), it would have been quite tiring for an individual participant to undergo pre- and post-ERP assessments, alongside equivalence training and testing, within a single experimental session. FAA, however, can be computed across fewer trials, permitting baseline and follow-up assessments. Second, ERP data can be difficult to interpret, given that the same electrophysiological components (e.g., the P300) can be employed to support a wide variety of interpretations (Patel & Azzam, 2005). FAAs on the other hand are almost exclusively employed as measures of affect, reducing the possibility for interpretational ambiguity (again, see Briesemeister et al., 2013). Third, since there were no dependent measures other than EEGs in the Amd et al. study, it is unknown to what extent the reported findings would have corroborated with other behavioral measures, which is why in the present experiment we employed self-reports and IATs alongside EEG as dependent measures. Fourth, since Amd et al. directly paired the A stimuli in each of the equivalence classes with emotion-eliciting imagery, a Pavlovian account of the reported effects cannot be completely obviated (cf. Tonneau, 2001). Consequently, the present experiment did not involve any direct pairings between emotion-eliciting stimuli and the stimuli employed during conditional discrimination training/testing. Finally, we examined ToF in accordance with comparative, rather than equivalence, relations, combining all of the aforementioned questions and measures into one grand study.

In the present experiment, participants were initially presented with five "face" stimuli (A, B, C, D, and E) multiple times in a randomized sequence. The participants had to rate how "happy" or "unhappy" a face made them feel on a 5-point Likert scale while their EEG was recorded. Next, they underwent an IAT similar to those employed in previous emotion ToF research (Amd & Barnes-Holmes, 2014; Amd & Roche, 2015). Specifically, participants were asked to categorize the B and D stimuli with happy and unhappy words alternately. Following acquisition of the self-reports, EEG, and IAT performances (i.e., pre-TI), participants underwent training and testing in order to establish two abstract shapes as *happier-than* and *unhappier-than* contextual cues, respectively. Participants were then trained on the conditional discriminations A+/B−, B+/C−, C+/D−, and D+/E− in the presence of the *happier-than* cue only, followed by tests for TI in the presence of both the *happier-than* and *unhappier-than* cues. Finally, participants rated the A, B, C, D, and E stimuli again as their EEG was recorded, followed by a second exposure to the IAT.

For participants who successfully demonstrated TI, we predicted the following: For the self-report, the happiness ratings for the face stimuli would correspond to a A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D$>_{happy}$ E hierarchy. More specifically, we predicted that A and B would be rated as more happy than during baseline, whereas D and E would be rated as less happy than during baseline. For the IATs, we predicted that B would be categorized more fluently with happy words, and D with unhappy words, than in their pre-TI performance, as previously demonstrated (Amd & Barnes-Holmes, 2014; Amd & Roche, 2015). For the EEG measure, it should be noted that no investigation has previously employed FAA measures to quantify derived emotional responses, or, more importantly, to parse out a series of such responses in relation to each other. Given that physiological measures are often inappropriate for quantifying relative response strengths across a range of stimuli greater than two, such measures tend to be confined to merely distinguishing which of *two* stimuli is more salient emotionally (cf. Boyle, Roche, Dymond, & Hermans, in press; Dougher, Hamilton, Fink, & Harrington, 2007). For this reason, our predictions regarding FAA measures of ToF were conservative. Specifically, we predicted that the A stimulus should elicit greater left frontal (and/or relatively less right frontal) alpha activity than the E stimulus in the post-TI measurement, given their positions as end-anchors in the A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E series. That is, because A and E were the "happiest" and "unhappiest" members within the series following ToF, their contrasting "derived" valences should elicit the greatest differential in FAA magnitudes.

Method

Participants

Twelve participants were recruited via personal invitation from Maynooth University; by "personal invitation," we refer to the recruitment method used by the first author, which involved standing outside the college gym and randomly approaching students leaving the premises to see whether they would be interested in taking part in a psychology experiment. None of the participants were familiar with TI, and they were unknown by the authors. Each participant received two Cadbury chocolate bars for their time. Three participants were excluded from the final analysis due to the absence of at least 30 s of artifact-free EEG data, leaving a final sample of five females and four males ($n = 9$; $M = 21.7$ years, $SD = 1.8$). All participants were right-handed and had normal or corrected-to-normal vision. Individuals interested in participating were orally interviewed to affirm the absence of confounding medical histories (i.e., they did not report using prescribed or recreational drugs or having epilepsy, depression, or schizophrenia) and to find out the participants' programs of study (none of them were psychology majors). The study was approved by the Maynooth University biomedical research ethics committee.

Materials

All tasks were programmed in the E-Prime 2.0 environment (Schneider, Eschman, & Zuccolotto, 2012). The face stimuli employed during the relational training and testing phases constituted male and female silhouettes with labels depicting stimulus "names" placed over the bottom half of each of the faces (see Fig. 1 for all of the stimuli used). The face stimuli will be referred to here as A, B, C, D, and E. The contextual cues comprised abstract shapes with no predicted preexperimental history.[1] Additionally, 18 face stimuli (six happy, six neutral, and six unhappy) were taken from the Radboud Faces Directory (Langner et al., 2010) for contextual cue training and testing.

Electrophysiological setup and analysis

EEG activity was recorded from 26 silver/silver chloride electrodes mounted in an elastic cap fastened with a chin strap (Easy-Cap, Herrsching, Germany). The amplifier used for recording was supplied by BrainVision (QuickAmp; BrainProducts GmbH, Germany). The electrode sites were determined according to the International 10–20 System for electrode placement. Specifically, EEG activity was recorded over frontal (F1, F2, F3, F4, F5, F6, F7, F8, Fz), central (C1, C2, C3, C4, Cz), temporal (T1, T2), parietal (P1, P2, P3, P4, P5, P6, Pz), and occipital (O1, O2, Oz) regions. Electroconductive gel (Abralyt 2000; Easy-Cap) was placed on each electrode site with a 10-ml flat-tipped syringe. The reference electrode during recording was located on the tip of the nose and was re-referenced to Cz during analysis.[2] Four additional electrooculography electrodes were positioned around the eyes in order to

Fig. 1 A, B, C, D, and E stimuli (top) were employed for training the conditional discriminations in the presence of Cue 1, followed by tests for mutual and combinatorial entailment in the presence of Cue 2 (middle row). Examples of the emotional faces employed during training of the contextual cue functions are presented near the bottom left. Following training, participants were presented with novel emotional face pairs in trials testing for the establishment of the *happier-than* and *unhappier-than* contextual cue functions (bottom right).

detect eye movements. Specifically, two electrodes were placed at the external canthi of the eyes (for recording horizontal movements), whereas two more were attached to the inferior and superior ridges of the left eye, respectively, to record vertical movements. Blinks were averaged offline, after which a blink reduction algorithm for automated artifact correction was applied to the data set (Ille, Patrick, & Scherg, 2002; Scherg & Berg, 1991). Ocular and muscular artifacts were minimized a priori through careful instructions (cf. Hagemann, 2004).

Impedances were kept between 5 and 9 kΩ. The voltage differences between the 26 electrodes and the reference electrode were extracted as electrical waveforms and amplified with a bandpass of 0.16–60 Hz and a gain of 1000. The conversion rate was set at 2000 Hz per channel within a 150-V range. The recordings were notch-filtered at 50 Hz. The EEG data were digitized at a sampling rate of 256 Hz per channel and were analyzed offline using BESA software. Epochs were defined from 0 to 2,048 ms post-stimulus-onset. EEG data were used to identify individual transition frequencies (TFs) and individual alpha frequencies (IAFs), allowing the definition of 5-Hz-wide individual alpha bands, which were calculated over anterior (F3, F4, F7, F8) electrode sites.[3] Fast

---

[1] Research by Kawahara and Shinohara (2012) suggested that "angular" shapes may be more emotionally loaded than "rounded" ones, in that the former appear to be associated with emotional attributes related to arousal rather than to valence per se (e.g., "shocked" rather than "sad"). In the present study, however, even if the more "angular" Cue 2 was emotionally "loaded" a priori, note that the conditional discriminations involving the A–B, B–C, C–D, and D–E stimulus pairs were reinforced in the presence of Cue 1 *only*.

[2] Identifying appropriate reference sites in alpha asymmetry research has been a contentious issue (Hagemann, 2004; Hagemann, Naumann, & Thayer, 2001), particularly if FAA is computed as the difference in separately derived power values of individual electrodes referenced to the vertex (Hagemann et al., 2001, p. 848). Although the details of this matter are beyond the scope of the present article, it suffices to say that Cz has been the most commonly used reference in the FAA research (Coan & Allen, 2003) and continues to be a popular reference of choice (cf. Huang et al., 2014). Given that numerous studies have successfully predicted relationships using the Cz reference, it may be the case that published results utilizing Cz may not only reflect an FAA effect in relation to the construct under investigation, but also sources of variance unique to Cz (Allen, Coan, & Nazarian, 2004). Given that the specific topography of the EEG response was not the variable of interest in the present study, the Cz reference was employed so that the present results may be compared with the wider literature on induced FAAs.
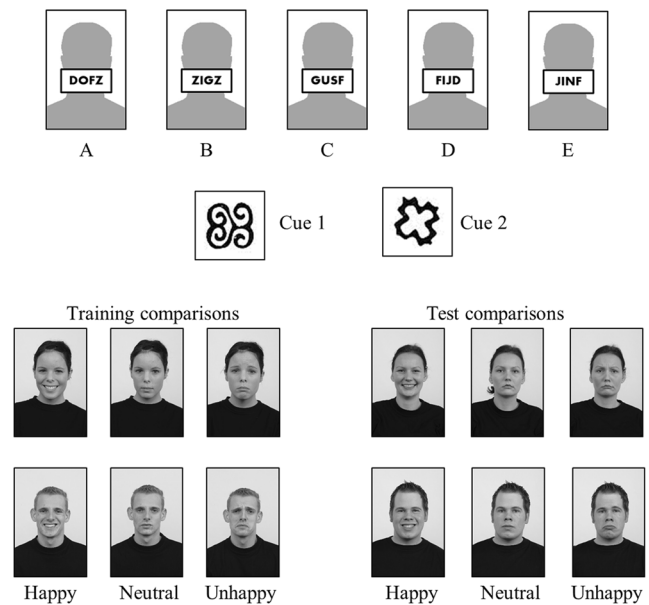
[3] Although individual alpha bands are typically calculated over posterior sites in order to minimize ocular "noise" (which can resemble alpha activity—see Hagemann, 2004, pp. 159–162), ancillary analyses of the aforementioned participants' data sets revealed no notable differences between frontal and posterior alpha peaks (SDs < 1 Hz).

Fourier transforms were conducted over these 5-Hz bands to calculate power values that were normalized and log-corrected (see the Procedure section for additional details). FAA values were calculated by subtracting the natural logs of (F8 + F4/2) from (F7 + F3/2), with positive values indicating greater left frontal activation and negative values greater right frontal activation (Harmon-Jones, Gable, & Peterson, 2010).

Procedure

Phase 1—Baseline assessment

Self-report/frontal alpha asymmetry The goal of this segment was to acquire happiness ratings for the A, B, C, D, and E stimuli along a 5-point Likert scale while EEG data were being collected. Participants were seated in front of a 14-in. computer screen in a Faraday-shielded room within a temperature range of 18 ± 2 deg Celsius. After attaching the electrodes and checking that the impedance levels were appropriate (i.e., <10 kΩ), participants were instructed to relax and look at a white fixation point on a blank, black screen for 3 min, followed by a photic prompt to close their eyes for a further 3 min, during which their EEG was recorded. The procedure of measuring eyes open followed by eyes closed allowed us to determine TFs and IAFs, which in turn allowed for defining individual alpha bands (Klimesch, 1999).

According to Klimesch (1999), computing individualized alpha bands and their subcomponents involves the following steps: First, calculate power spectra separately for the eyes-closed and eyes-open conditions. Second, mark the frequency at which "theta" (operationally defined as the enhanced spectra during task engagement) and "alpha" (operationally defined as the suppressed spectra during task engagement) intersect as the *transition frequency*, or TF (the TF marks the lower tail of an individual's alpha range) during the eyes-open condition. The TFs for individuals participating in the present experiment were 6.5 Hz (for P2, P3, and P6), 7 Hz (for P1), 7.5 Hz (for P4), and 8 Hz (for P5, P7, P8, and P9). Third, identify the frequency peak after the TF and deem this the participant's individual alpha peak frequency. The peak frequencies fell within a range of 10–11.5 Hz for all participants in the present study, as would be expected for healthy young adults (Bazanova & Vernon, 2014, p. 8). Finally, one may compute the "upper alpha" (UA) band using the formula $5 - [IAF - TF]$. Note that because we did not predict variations in the subcomponents of the alpha band, identifying the individual UA bandwidths was not deemed necessary. Presently, individual alpha bands were simply computed as the range between TF and TF + 5, given that the (alpha) frequency range between the TF and UA markers is approximately 5 Hz wide (Klimesch, 1999, p. 171).

Following eyes-open/eyes-closed recording, the participant was instructed to look at the computer screen as the following instructions appeared

> Welcome. In this part of the experiment, you will first see an image come up on screen. Please pay attention to the image. After a few moments, five boxes will appear near the bottom of the image. Please indicate how "happy" or "unhappy" the image makes you feel by clicking on the appropriate box (1 = *happy*, 2 = *somewhat happy*, 3 = *neutral*, 4 = *somewhat unhappy*, 5 = *unhappy*) with the mouse. Please respond using the index finger on your right hand ONLY during all stages of the experiment. IT IS IMPORTANT that you refrain from moving your head and blinking/moving your eyes when the image first appears—you may blink/move your head after the boxes have appeared. Please ask the experimenter if you have any further questions—otherwise press any key to begin. . . .

Upon pressing any key, a blank gray screen with a white fixation cross appeared for 300 ms, followed by a face stimulus (either A, B, C, D, or E) in the center of the screen. After 2,000 ms, the mouse cursor and five white boxes containing one of the numbers 1, 2, 3, 4, and 5 printed in black appeared near the bottom half of the screen. Within an unlimited response window, participants had to move the cursor to one of the boxes and click on it to continue. Clicking a box switched its background from white to yellow for 150 ms. This was followed by an intertrial interval (ITI) of 300 ms, consisting of a fixation point on a gray screen, followed by the next trial. The A, B, C, D, and E stimuli appeared 15 times each for a block total of 75 trials (see Fig. 2). Completion of 75 trials terminated the phase.

Implicit association test Participants were next exposed to IATs similar to the ones employed by Amd and Barnes-Holmes (2014). Participants were presented with the following instructions on the screen:

> Welcome. You will now be presented with a set of words and faces to classify into groups using the "z" and "m" keys on the keyboard. Please place your left and right index fingers on the "z" and "m" keys. You have to 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 . . .

Upon pressing a key, the labels ZIGZ and FIJD appeared in a blue font on the left and right sides, respectively, of the top
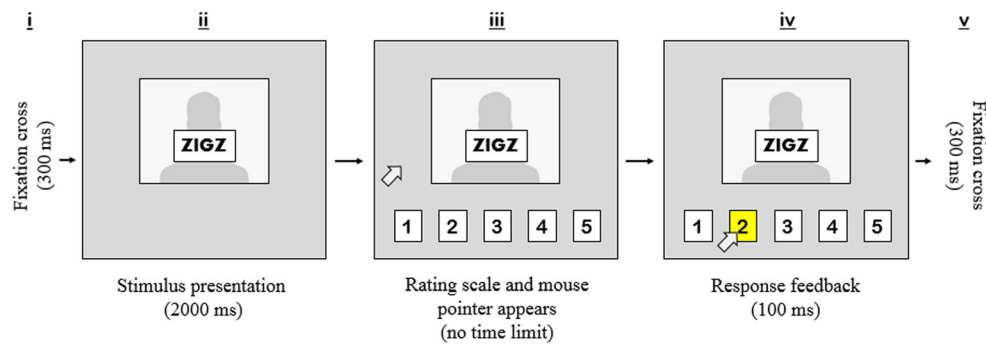
Fig. 2 Illustration of the trial sequence (i–v) during the rating/ electroencephalographic-recording task in Phase 1. Participants first viewed a fixation point on a blank, gray screen for 300 ms (i), followed by the presentation of an A, B, C, D, or E stimulus. The image remained on the screen for 2,000 ms (ii), after which five numbered boxes appeared near the bottom of the screen, along with a mouse cursor (iii). The participant had to select one of the boxes in order to complete the trial. When a participant clicked on a box, it turned yellow for 100 ms (iv), after which the fixation point reappeared (v), signaling the end of the trial. Participants were exposed to 75 trials involving 15 presentations each of the A, B, C, D, and E stimuli.

half of a black screen (see Fig. 3). Note that ZIGZ and FIJD were the labels for the stimuli B and D (see Fig. 1). In each trial, either stimulus B or D appeared in the center of the screen for 3,000 ms. If stimulus B appeared, the correct response was the letter "z"; if stimulus D appeared, the correct response was the letter "m." Emitting a correct response was followed by the word "Correct!" presented in green font for 500 ms. An incorrect response was followed by a red "X" for the same duration. If no response was detected for 3,000 ms, a message stating "Too slow!" appeared in gray font for 500 ms. Feedback was followed by the onset of a blank ITI for 200 ms, followed by the next trial.

Participants underwent 40 trials in this fashion, in what was deemed Block 1. In Block 2, the ZIGZ and FIJD labels disappeared, replaced by the labels HAPPY and UNHAPPY at the top left and right sides of the screen, respectively, in yellow font. At this stage, happy ("pleasant," "overjoyed," "delighted," "glad") or unhappy ("sad," "miserable," "gloomy," "unpleasant") words individually appeared in a randomized sequence near the center of the screen. If a happy word appeared, the correct response was "z," and if an unhappy word appeared, the correct response was "m." After 40 trials in Block 2, participants progressed to Block 3, which involved presenting the labels ZIGZ/HAPPY and FIJD/UNHAPPY on the left and right sides of the screen, respectively. Block 3 involved the presentation of stimulus B, stimulus D, a happy word, or an unhappy word in the center of the screen. If either stimulus B or a happy word appeared, the correct response was "z"; if stimulus D or an unhappy word appeared, the correct response was "m." Block 3 comprised 20 trials. Block 4 was an extension of Block 3, but with 40 trials.

In Block 5, the HAPPY and UNHAPPY labels were removed, and the FIJD and ZIGZ labels switched their left/ right positions. That is, either stimulus B or stimulus D appeared near the center of the screen, where the correct response in the presence of stimulus B was "m," and the correct response in the presence of stimulus D was "z." Block 5 consisted of 20 trials. In Block 6, the HAPPY and UNHAPPY labels reappeared at the left and right sides of the screen, respectively, with the FIJD and ZIGZ labels still being presented in their previous positions. Block 6 involved the presentation of stimulus B, stimulus D, a happy word, or an unhappy word near the screen center. If either stimulus D or a happy word appeared, the correct response was "z"; if stimulus B or an unhappy word appeared, the correct response was "m." Block 6 comprised 20 trials. Block 7 was an extension of Block 6, but with 40 trials. The completion of Block 7 marked the end of the pre-TI (baseline) IAT. It should be noted that the block sequence was counterbalanced between participants.

Phase 2—Relational training and testing

There were two goals of the present phase: first, to establish two arbitrary shapes as contextual cues functionally equivalent to the phrases "happier than" and "unhappier than," respectively; second, to establish a five-member A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E hierarchy. The phase was divided into two segments.

Segment 1—Establishing the contextual cues The goal of Segment 1 was to establish Cue 1 and Cue 2 as functionally equivalent to the English phrases "happier than" and "unhappier than," respectively (cf. Amd & Roche, 2015). The segment commenced with the following instructions presented on screen:

> Welcome. During this part of the experiment, you will see an image appear on the top half of the screen, followed by two more images near the bottom half of the screen. You will have to select one of the two images near the bottom of the screen in order to continue. To
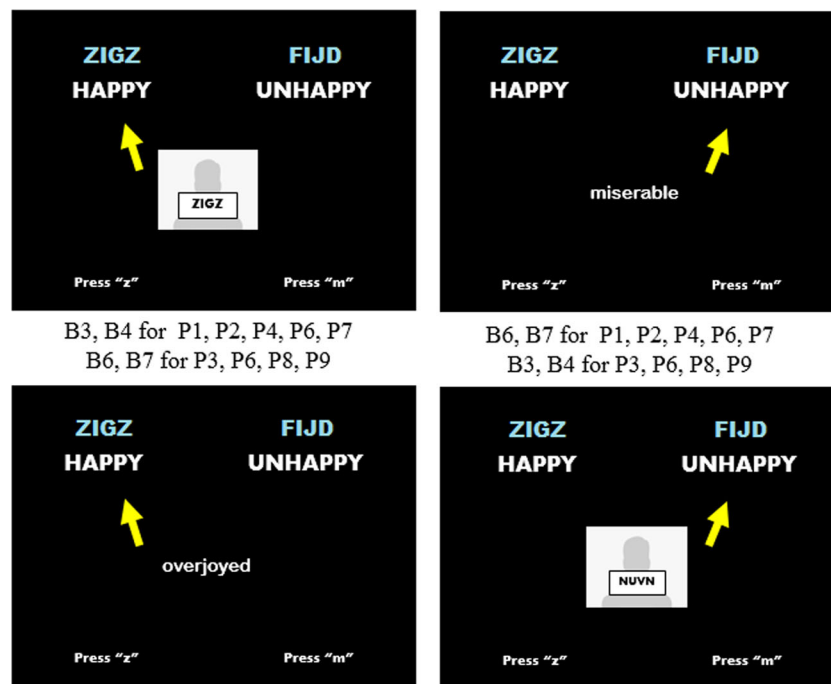
Fig. 3 Illustrations of implicit association test Trial Blocks 3, 4, 6, and 7 (B3, B4, B6, and B7). Participants 1, 2, 4, 6, and 7 (P1, P2, P4, P6, and P7) were exposed to the trial types on the left side of the figure during B3 and B4, followed by the trial types on the right side of the figure during B6 and B7. Alternatively, P3, P6, P8, and P9 were first exposed to the trial types on the right side of the figure during B3 and B4, followed by the trial types on the left during B6 and B7.

select the stimulus on the left, press "z"; to select the stimulus on the right, press "m." You will receive feedback for your selections. Try to make as many "correct" selections as you can! You will make some mistakes at first and that is okay. You will learn as you progress. Please ask the experimenter if you have any questions. Otherwise, press any key to begin.

Upon pressing a key, a blank, gray screen appeared for 500 ms, followed by the appearance of either Cue 1 or Cue 2 at the top of the screen for 2,000 ms. Next, two face stimuli appeared near the bottom half of the screen. The emotional expressions of the presented face pairs were as follows: happy—neutral, neutral–unhappy, and happy–unhappyhappy/neutral, neutral/ unhappy and happy/unhappy (the left and right positions of each face were counterbalanced across trials). All stimuli remained on screen until the participant had emitted an appropriate response (i.e., pressing "z" or "m" to select the face on the left or right, respectively). Corrective feedback was provided by the onscreen message "Correct!" in green font, or the word "Wrong" in red font, on a gray screen for 1, 000 ms. A blank 500-ms ITI followed, succeeded by the subsequent trial. Participants were thus instrumentally conditioned to select the relatively happier face in the presence of Cue 1, or the unhappier face in the presence of Cue 2. The position and presentation of the

comparison pairs were counterbalanced across trials for each participant. All participants underwent training until they had produced ten consecutive correct responses, after which the following message appeared

You will now continue the same procedure without any feedback. Please press any key when ready.

Participants were once again presented with either Cue 1 or Cue 2 near the top of the screen, followed by a novel pair of emotional faces. They had to press either the "z" or the "m" key to select a comparison in order to progress to the next trial. Responses were followed by a 1,500-ms ITI consisting of a blank gray screen, after which the following trial commenced. The completion criterion for the 30-trial test block was the production of ten consecutive correct discriminations in the presence of novel face pairs (see Fig. 1, bottom right panel). Meeting criterion was followed by the message "Please contact the experimenter" appearing on the computer screen, signaling the end of the contextual cue training and testing segment.

Establishing the A > B > C > D > E series Following the emission of ten correct discriminations in the absence of any corrective feedback, the following message appeared on the participant's screen.

By now you have an idea of how the procedure works. You should continue to respond as you have so far (press "z" for left; press "m" for right). You will receive corrective feedback during the beginning, but sometimes you will not. Please pay attention to what you learn throughout the task. Please ask the experimenter if you have any questions, or press any key to begin.

Following a keypress, Cue 1 appeared near the top of the screen for 2,000 ms. Next, one of four comparison pairs (B/C, A/B, C/D or D/E) appeared near the bottom half of the screen. Note that the presentation and positioning of the comparison pairs were randomized across trials. Participants had to select a comparison from the presented pair in order to continue. Responses were followed by corrective feedback (the words "Correct!" or "Wrong"). Participants were trained on four conditional discriminations: B+/C–, A+/B–, C+/D–, and D+/E– (in no specific order), in the presence of Cue 1 (i.e., *happier-than*) only. Participants were exposed to each comparison pair ten times in a random order across 40 trial training cycles until they had produced ten consecutive correct responses. Once this criterion was met, the computer program began to fade in test trials for functional symmetry (50% of trials), while reducing the proportion of feedback for training discriminations from 100% to 50%. Note that no corrective feedback was provided during test trials.

Tests for symmetry involved presenting the comparison pairs shown during training trials, but in the presence of Cue 2 (*unhappier-than*) as the sample. For example, if participants had been rewarded for choosing A over B (A+/B–) in the presence of Cue 1 during training, a successful demonstration of symmetry would involve the participant choosing B over A (B+/A–) in the presence of Cue 2. Participants were exposed to 16 trials testing for symmetry and 16 trials testing for the relations trained previously. This was followed by tests for functional transitivity, which involved exposure to the A/C, B/D, C/E, B/E, A/E and A/D comparison pairs in the presence of Cues 1 and 2 equal numbers of times across a 24-trial block. Completion of the 24 test trials was followed by a message stating "Please wait for the experimenter," signaling the completion of Phase 2.

Phase 3—Post assessment

Phase 3 involved a readministration of all of the tasks from Phase 1. That is, participants had to rate the A, B, C, D, and E stimuli 15 times again over a 75-trial block while their EEG was recorded. This was followed by a readministration of the previous IAT in which participants were required to indirectly relate the B and D stimuli with happy and unhappy words alternately. Completion of the IAT occasioned the onscreen message "Thank you," ending the experiment.

Results

Relational training and testing (Phase 2)

Recall that the first goal of Phase 2 was to establish Cue 1 and Cue 2 as meaning *happier-than* and *unhappier-than*, respectively. All participants met the training criterion of emitting ten consecutive correct responses within 50 training trials. That is, they consistently selected the relatively "happier" comparison in the presence of Cue 1, and the "unhappier" comparison in the presence of Cue 2. These performances remained robust (100% accuracy) during the ensuing test trials with novel face pairs, in which participants continued selecting the happier face in the presence of Cue 1, and the unhappier face in the presence of Cue 2.

The second goal of Phase 2 was to establish an A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E hierarchy. All participants produced ten consecutive correct conditional discriminations during training. That is, they responded to the B+/C-, A+/B-, C+/D- and D+/E- discriminations in the presence of Cue 1 ten times consecutively. Specifically, participants P3, P4, P8, and P9 met the criterion within one 40-trial block; P2 and P7 met the criterion within two 40-trial blocks; and P1, P5, and P6 met the criterion within three 40-trial blocks. All participants underwent the same numbers of testing trials (16 trials for the relations directly trained, 16 for symmetry, and 24 for transitivity). For test trials involving relations directly trained, P1 and P5 produced accuracies of 88% and 94%, respectively, whereas everyone else responded with 100% accuracy. In contrast, during the tests for symmetry, P1, P5, and P6 produced accuracies of 69%, 13%, and 75%, respectively, whereas the remaining participants responded with 100% accuracy. In tests for transitivity, P1, P5, and P6 produced accuracies of 46%, 29%, and 58%, respectively. P3, P4, P7, and P8 produced accuracies of 92%, and P2 and P9 of 100% (see Table 1 for details).

In summary, P2, P3, P4, P7, P8, and P9 "passed" the tests for both symmetry and transitivity, indicating that the A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E hierarchy could be inferred to have been established for these participants. The same could not be said for P1, P5, and P6, given their performances (see Table 1). Nevertheless, the data of P1, P5, and P6 have been retained to serve as a contrast to the performances of the other participants.

Self-report (Phase 1 vs. Phase 3)

Averaged ratings of the A, B, C, D, and E stimuli have been presented for individual participants in Table 2. The values were averaged from the 15 ratings made per stimulus per condition (pre-TI vs. post-TI). The delta values in Table 2 indicate changes in ratings from the pre-TI to the post-TI measures. A positive delta value indicates that the post-TI ratings

Table 1 Individual performances for test trials for directly trained, symmetry, and transitive relations

Count[*]	Cue[**]	Pairs[**]	P1F[+]	P2	P3	P4	P5F	P6F	P7	P8	P9
Conditional Discrimination Training[***]											
<30	Cue 1	A–B	26	14	8	9	27	18	13	8	8
<30	Cue 1	B–C	24	18	9	7	18	26	17	8	9
<30	Cue 1	C–D	19	17	8	8	12	27	18	9	7
<30	Cue 1	D–E	21	16	8	8	22	23	17	9	8
Tests for Trained Discriminations[++]											
4	Cue 1	A–B	75%	100%	100%	100%	75%	100%	100%	100%	100%
4	Cue 1	B–C	100%	100%	100%	100%	100%	100%	100%	100%	100%
4	Cue 1	C–D	100%	100%	100%	100%	100%	100%	100%	100%	100%
4	Cue 1	D–E	75%	100%	100%	100%	100%	100%	100%	100%	100%
	Average % Correct		*88%*	*100%*	*100%*	*100%*	*94%*	*100%*	*100%*	*100%*	*100%*
Tests for Symmetry											
4	Cue 2	B–A	75%	100%	100%	100%	0%	100%	100%	100%	100%
4	Cue 2	C–B	50%	100%	100%	100%	25%	50%	100%	100%	100%
4	Cue 2	D–C	100%	100%	100%	100%	25%	50%	100%	100%	100%
4	Cue 2	E–D	50%	100%	100%	100%	0%	100%	100%	100%	100%
	Average % Correct		*69%*	*100%*	*100%*	*100%*	*13%*	*75%*	*100%*	*100%*	*100%*
Tests for Transitivity											
2	Cue 1	A–C	0%	100%	100%	100%	50%	0%	100%	100%	100%
2	Cue 1	B–D	50%	100%	100%	100%	50%	100%	50%	50%	100%
2	Cue 1	C–E	50%	100%	100%	100%	0%	50%	100%	100%	100%
2	Cue 1	B–E	100%	100%	100%	100%	0%	100%	100%	100%	100%
2	Cue 1	A–D	50%	100%	100%	100%	50%	100%	100%	100%	100%
2	Cue 1	A–E	100%	100%	100%	100%	50%	100%	100%	100%	100%
2	Cue 2	A–C	50%	100%	100%	100%	50%	0%	100%	100%	100%
2	Cue 2	B–D	0%	100%	50%	50%	0%	50%	50%	50%	100%
2	Cue 2	C–E	0%	100%	100%	100%	0%	0%	100%	100%	100%
2	Cue 2	B–E	0%	100%	50%	100%	0%	0%	100%	100%	100%
2	Cue 2	A–D	50%	100%	100%	50%	50%	100%	100%	100%	100%
2	Cue 2	A–E	100%	100%	100%	100%	50%	100%	100%	100%	100%
	Average % Correct		*46%*	*100%*	*92%*	*92%*	*29%*	*58%*	*92%*	*92%*	*100%*

[*] Total number of trials per comparison pair per cue type. Note that during training, no comparison pair was presented more than 30 times for any given participant. [**] The cue type and the corresponding comparison pair. [***] Number of exposures to each training-trial type. The criterion during conditional discrimination was. [+] "P" designates the participant ID; "F" designates participants who did not demonstrate TI. So, for instance, "P1F" refers to Participant 1, who failed to demonstrate TI, and "P2" is Participant 2, who did demonstrate TI. Each column depicts the numbers of correct responses made. Note that the criterion for training (first four rows) was ten consecutive correct responses, whereas the testing criterion was set at eighteen correct responses out of twenty trials. [++] Percentages of correct responses per participant per condition during the test trials.

were "happier" than the pre-TI performance—a negative delta value indicates the opposite (i.e., post-TI ratings were "unhappier"). The participants who demonstrated TI produced positive deltas for stimuli A and B, and negative deltas for stimuli D and E (see Table 2).

Figure 4 illustrates the rating trends across individual participants. For the participants who demonstrated TI, the happiness ratings were highest for A and decreased incrementally from A to E, although there was individual variability (e.g., P8 rated stimulus C as less happy than stimulus D, and P9 rated stimulus C as happier than stimulus B or D). In contrast, among those who failed to show TI, the trends appeared less robust. Specifically, P1's ratings across the five stimuli did not significantly differ from pre-TI to post-TI, whereas P6 showed a trend in the opposite direction. P5 appears to show a rating trend similar to those of the participants who demonstrated TI, although the trend was skewed by the high rating provided for stimulus A.

Statistical analyses were conducted for participants who passed the tests for symmetry and transitivity. Specifically, a

Learn Behav (2016) 44:175–190

Table 2 Baseline and post ratings along a 5-point scale (5 = *happy*, 4 = *somewhat happy*, 3 = *neutral*, 2 = *somewhat unhappy*, 1 = *unhappy*)

ID[*]	A-BAS	A-POST	ΔA	B-BAS	B-POST	ΔB	C-BAS	C-POS	ΔC
P1F	4.8	5	(0.2)[**]	4.6	4	(–0.6)	3.4	3.8	(0.4)
P2	3.4	5	(1.6)	1.6	4	(2.4)	2.6	2.4	(–0.2)
P3	3.4	5	(1.6)	2.8	5	(2.2)	3	2.4	(–0.6)
P4	2.4	5	(2.6)	3	4.2	(1.2)	3	3.8	(0.8)
P5F	4	4.2	(0.2)	2.4	2	(–0.4)	2.4	2.6	(0.2)
P6F	2	2	(0)	3.8	2.4	(–1.4)	2.6	4	(1.4)
P7	3	5	(2)	4	5	(1)	4	4	(0)
P8	4	5	(1)	3	5	(2)	2.8	2.4	(–0.4)
P9	3.4	5	(1.6)	3.6	3.2	(–0.4)	5	4.2	(–0.8)

ID	D-BAS	D-POST	ΔD	E-BAS	E-POST	ΔE
P1F	3.8	5	(0.2)	4	4.6	(0.6)
P2	2.4	2.2	(–0.2)	3.2	2	(–1.2)
P3	3.8	2	(–1.8)	4.8	2	(–2.8)
P4	2.6	2	(–0.6)	3	1	(–2)
P5F	2.6	2.2	(–0.4)	2.8	2.8	(0)
P6F	4	2	(–2)	3.6	5	(1.4)
P7	3	2.4	(–0.6)	3.4	1.2	(–2.2)
P8	3.4	2.8	(–0.6)	5	1.8	(–3.2)
P9	4.2	2.4	(–1.8)	4.2	2	(–2.2)

[*] The IDs of all nine participants who completed the procedure. Participants with an "F" (i.e., P1F, PF5, P6F) did not meet criterion in the tests for entailment; briefly, the "F" indicates "Fail." [**] Delta (Δ) columns show changes in ratings (post – baseline). Positive delta values indicate that a stimulus was rated happier than at baseline; negative delta values (underlined) indicate that a stimulus was rated unhappier than at baseline. For example, P9 rated stimulus A as 3.4 during baseline and 5 during the post ratings, so her happiness rating for A increased by 1.6 (= ΔA).

paired one-way analysis of variance (ANOVA) showed a main effect of stimulus type on happiness ratings, $F(5, 14) = 44.39$, $p < .0001$, $\eta^2 = .69$. Post-hoc analyses using Tukey's HSD indicated that the A stimulus occasioned significantly happier ratings than the C ($p < .01$), D ($p < .01$), and E ($p < .01$) stimuli. Stimulus B occasioned significantly happier ratings than C ($p < .01$), D ($p < .01$), and E ($p < .01$). All other comparisons were nonsignificant.

Implicit association tests (Phase 1 vs. Phase 3)

Statistical analyses were conducted for IAT blocks B3, B4, B6, and B7, in accordance with the C4 algorithm outlined by Greenwald and colleagues (2003). No data were excluded from the analysis, since all recorded latencies were between 300 and 10,000 ms. Our analysis of the IAT data constituted the following steps: First, the mean latencies of "correct" responses 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. Second, the latencies for incorrect responses were replaced with the block means of correct-response latencies, increased by 600 ms. Third, the adjusted latencies were averaged for each block (B3μ, B4μ, B6μ, B7μ). Fourth, 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 within $+2 < d < -2$, which signifies the magnitude of the IAT effect.

Let us first examine the IAT performance of the participants who did not demonstrate TI (see Fig. 5 and Table 3). Whereas P1 did pair B with happy words more fluently after than during pre-TI, the difference was not large ($\Delta d = 0.04$). P5 and P6 produced performances orthogonal to that of P1 by pairing D with happy words more fluently, relative to their pre-TI performances. In contrast, all participants who demonstrated TI yielded positive Δd values, indicating that B was paired more fluently with happy words relative to pre-TI. This indicates that B was now implicitly "happier" than D.

Frontal alpha asymmetry

EEG was recorded for 6 min (3 min eyes open + 3 min eyes closed) prior to the onset of the stimulus rating task in Phase 1. Note that the individual alpha frequency bands computed ranged from 6.5 to 11.5 Hz (for P2, P3, and P6), 7 to 12 Hz (for P1), 7.5 to 12.5 Hz (for P4), and 8 to 13 Hz (for P5, P7, P8, and P9). FAA scores were computed by subtracting the natural log-α of the left hemisphere (F3 + F7/2) from the right (F4 + F8/2). Positive FAAs indicate relatively greater left prefrontal activity, whereas negative FAAs values indicate
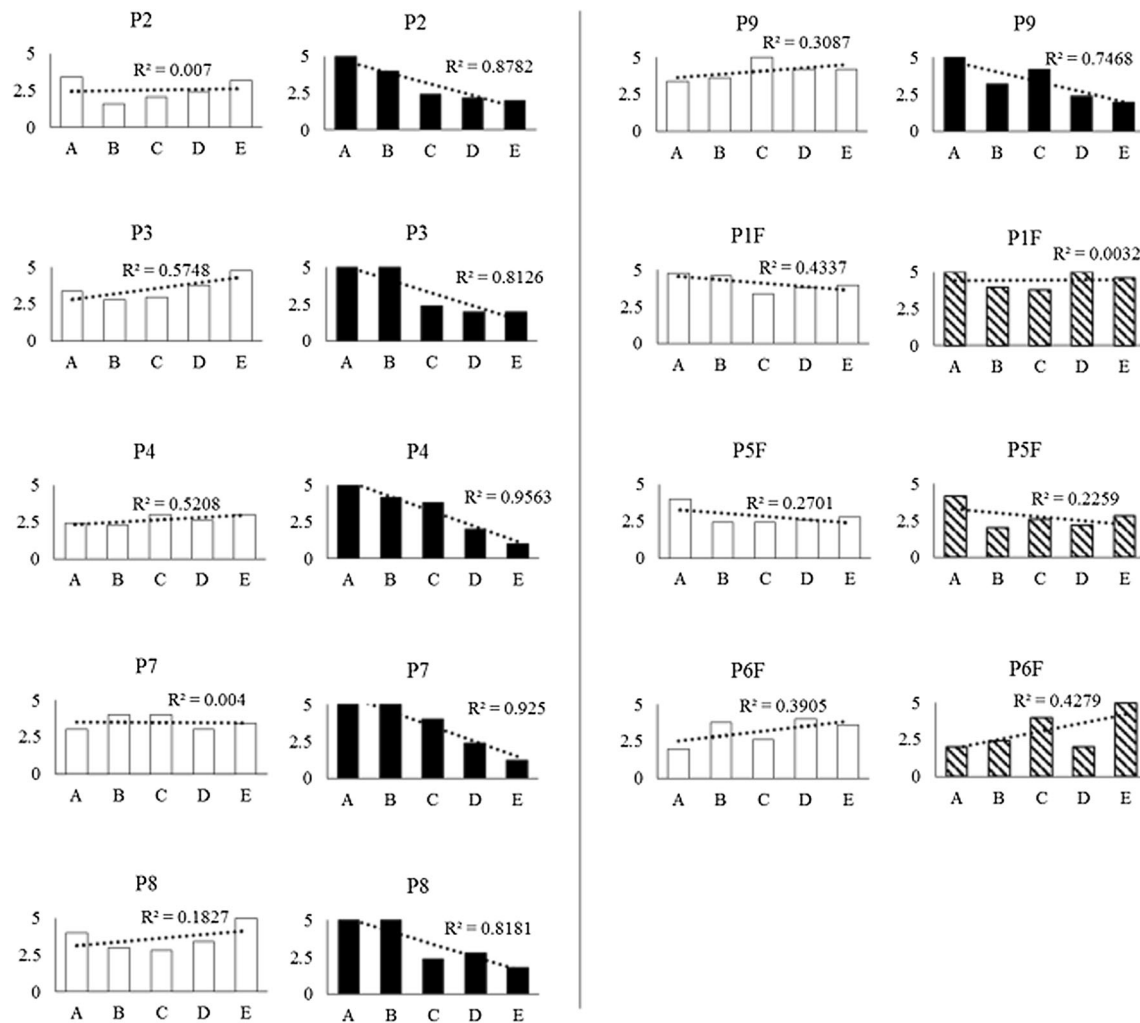
Springer

Fig. 4 Baseline and post-transitive-inference self-reports of the "happiness" level of the stimuli along a 5-point Likert scale (the *y*-axis) for the stimuli A, B, C, D, and E (along the *x*-axis). For the participants who passed both tests of entailment in Phase 2 (i.e., P2, P3, P4, P7, P8, and P9), baseline assessments are displayed in white bars, and post assessments are displayed in black bars. Note that the rightward trends (min. $R^2 > .745$) for these participants are in support of an A > B > C > D > E hierarchy. The post ratings for participants who did not pass the tests for entailment are represented in bars with diagonal lines.

relatively greater right prefrontal activity. The individual participants' FAA values acquired during the pre-TI and post-TI assessments are illustrated in Fig. 6. Statistical analyses were conducted for the participants who demonstrated TI. Specifically, a paired one-way ANOVA showed a main effect of stimulus type on FAAs, $F(5, 14) = 3.43$, $p < .02$, $\eta^2 = .15$. Post-hoc analyses using Tukey's HSD indicated that only stimulus A elicited a significantly greater FAA than did E ($p < .05$). All other comparisons were statistically nonsignificant ($p > .05$).

Upon visual inspection of the data, we can note that for P2, P3, P4, and P8, the FAAs elicited by the B stimulus were more positive than those elicited by the D stimulus. Conversely, for P7 and P8, greater FAAs were elicited by the D stimulus, relative to the B stimulus. For participants who failed to demonstrate TI, the FAAs elicited for the five stimuli did not appear to deviate visually from baseline, indicating that conditional discrimination training and testing had little effect on those who did not demonstrate TI.

Discussion

The present study involved establishing a five-member relational series, A > B > C > D > E, where ">" may be deemed as functionally equivalent to the phrase *happier-than*. Both previous to (pre-TI) and following (post-TI) training and tests for TI, participants were required to rate the "happiness" levels of the A, B, C, D, and E stimuli on a 5-point Likert scale while their EEG data were recorded. Participants next had to pair the B and D stimuli with happy and unhappy words alternatively in an IAT, again during pre-TI and post-TI performances. Among the participants who demonstrated TI, the self-report, IAT, and EEG measures demonstrated a transformation of
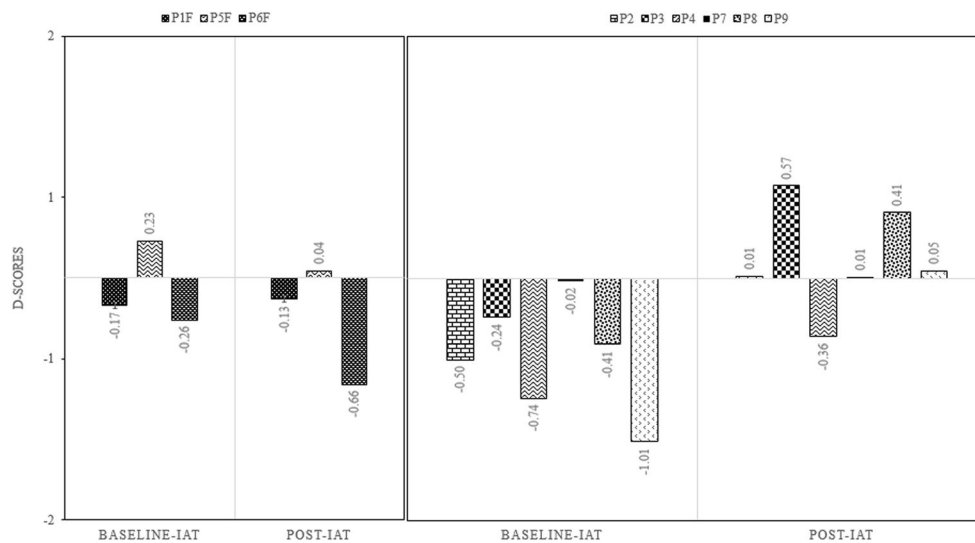
Fig. 5 Baseline and post-transitive-inference implicit association test (IAT) performances for P1, P5, and P6, who did not pass the tests for entailment, are presented in the left panel, and the performances for P2, P3, P4, P7, P8, and P9, who did pass both tests of entailment, are presented in the right panel. All of the participants who demonstrated entailment produced the predicted IAT performance by pairing B with happy words more fluently than at baseline.

"happiness" functions in the direction predicted. That is, the A and B stimuli were rated as *happier than* the D and E stimuli in the post-TI self-reports. Similarly, the B stimulus was categorized with happy words more fluently than with the D stimulus in the post-TI IAT. Finally, stimulus A elicited significantly greater FAAs than did stimulus E. These results indicate that the happiness functions of the stimuli had transformed in accordance with the structure of the A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E series. No published study has attempted to use self-reports in conjunction with performance-based (IAT) and spectral electrophysiological (FAA) measures, rendering the present findings both novel and informative.

A number of features of the present study warrant mention. First, although the present study originated from a relational frame theory (RFT) viewpoint, the results acquired may be accommodated with other behavioral accounts of TI, such as value transfer theory, or VTT (Steirn, Weaver, & Zentall, 1995; von Fersen, Wynne, Delius, & Staddon, 1991). According to Steirn and colleagues, value transfer takes place when "the member of each stimulus pair associated with nonreinforced responding acquires secondary positive value from the positive member of the pair" (p. 77). So, within the context of the A $>_{happy}$ B $>_{happy}$ C $>_{happy}$ D $>_{happy}$ E series established at present, VTT predicts that the "value" of stimulus A (the selection of which was never nonreinforced during training) should be highest, while that of stimulus E (the selection of which was never reinforced during training) should be lowest, as could be observed from the self-report and FAA data. VTT also predicts that the value of stimulus B (which had always been associated with A during training) would be greater than that of stimulus D (which had always

been associated with E during training), as could be deduced from the IAT data. If we were to substitute "happiness functions" for "value," the parallels between the predictions made by both the VTT and RFT accounts are clear.

Although there are obvious similarities between the VTT and RFT accounts, it is unknown whether verbally presented feedback in the form of the presented word "Correct!" may be sufficient to create associations analogous to those inferred from classical conditioning procedures used in non-human VTT experiments (see Weaver, Steirn, & Zentall, 1997). On the other hand, the processes by which such verbally presented feedback produces relations among stimuli is usefully described in the RFT account (i.e., in terms of the generalized reinforcing properties of some feedback words, and the generalization of association formation under the control of such feedback; see O'Hora, Barnes-Holmes, & Stewart, 2014). A purely VTT approach also faces difficulty when accounting for the results observed for participants who did *not* demonstrate TI. For example, when looking at the self-report data of P6 (Fig. 4) and/or the FAA data of P1 and P6 (Fig. 6), one may note that neither was stimulus A rated as the "happiest," nor did it elicit the greatest FAA, even though the selection of A was always reinforced for all participants. In effect, given that P1 and P6 did not demonstrate TI, one could argue that VTT is difficult to apply in this case. Recall, however, it should be noted that VTT is but one among numerous models accounting for TI, some of which acknowledge that TI may not always follow successful acquisition of baseline discriminations (cf. Delius & Siemann, 1998, and Wynne, 1995). Although a discussion of these various models is beyond the scope of the present article (but see Vasconcelos, 2008), it suffices to say at this stage that RFT remains silent with regard

Table 3 Baseline and post-transitive-inference performances on the implicit association test (IAT)

Baseline IAT Performance

ID	B3 + B4[*]	B6 + B7[*]	Pooled SD1[**]	d Score[+]	
P1F	747	666	476	–0.17	
P2	834	597	470	–0.50	
P3	619	544	311	–0.24	
P4	1073	567	681	–0.74	
P5F	687	809	530	0.23	
P6F	897	806	352	–0.26	
P7	601	590	536	–0.02	
P8	821	621	492	–0.41	
P9	1005	601	399	–1.01	

Post IAT Performance

ID	B3 + B4	B6 + B7	Pooled SD2	d Score	Δd[++]
P1F	620	580	307	–0.13	0.04
P2	747	753	481	0.01	0.51
P3	663	872	364	0.57	0.81
P4	842	620	616	–0.36	0.38
P5F	687	702	340	0.04	–0.19
P6F	824	512	473	–0.66	–0.40
P7	629	635	854	0.01	0.03
P8	608	811	496	0.41	0.82
P9	595	612	349	0.05	1.06

A positive d score indicates that B is happier than D; a negative d score indicates that D is happier than B. [*] Corrected means averaged for blocks B3 and B4 (B3 + B4) and for blocks B6 and B7 (B6 + B7). [**] SD1 is the standard deviation pooled from B3 and B6; SD2 is the standard deviation pooled from B4 and B7. [+] Difference scores indicate the magnitude of the IAT effect. A positive d score indicates B $>_{happy}$ D; a negative d score indicates D $>_{happy}$ B. [++] Outlines the change in d scores relative to baseline. Note that only P5 and P6 paired D with happy words more fluently relative to baseline, as indicated by the negative Δd values. All other participants paired stimulus B with happy words more fluently than at baseline, as indicated by the positive Δd values.

to the performances of P1, P5, and P6. In any case, the significant FAA effects observed between A and E, and the differences in self-report responses recorded across A, B, D, and E, may be accounted for almost equally well by both RFT and VTT.

A final feature worth noting is the demonstrated sensitivity of FAAs to *derived* emotional effects. As a neurophysiological metric of approach/avoidance motivations (Coan & Allen, 2004), FAAs have been typically observed following the presentation of emotion-eliciting stimuli (Davidson, 2004). In the present study, however, both the FAAs and self-report data illustrate significant differences between A/B and D/E following tests for TI. Yet, when viewing the baseline ratings/FAAs, no similar differences between A/B and D/E could be observed. Given that no other published study has employed individualized FAAs to assess for derived emotional effects,

the present findings are promising with respect to a physiological metric for assessing emotional ToFs.

Of course, the present study has some limitations worth noting. First, the number of participants may appear quite low ($n = 10$), particularly when considering that only six participants demonstrated TI. On balance, the effect sizes reported for both the self-report ($\eta^2 = .69$) and FAA ($\eta^2 = .15$) metrics are quite robust. Additionally, it may interest the reader to note that meeting six predictions (four for the self-report, one for the IAT, and one for the FAA) in five participants (Tables 1, 2, and 3) is less likely to be attributable to random chance than 29/30 participants meeting a single prediction (given that the conditional probabilities of all predictions are held constant).[4] In any case, although future research should replicate the present findings with an increased sample, with the aforementioned point we hope to convey that having a small sample size can be probabilistically mitigated via increasing the number of experimental predictions made.

A second limitation of the present study is that only six out of ten participants demonstrated TI, even after meeting the criterion during training trials (see Table 1). Failure to exhibit TI following acquisition of the baseline relations may simply reflect a lack of task awareness of the A > B > C > D > E hierarchy (Lazareva & Wasserman, 2010; Martin & Alsop, 2004). Specifically, Martin and Alsop reported that participants who failed to demonstrate TI despite successfully undergoing acquisition were simply less "aware" of the structure of the relational hierarchy trained (A > B > C > D > E in the present experiment). Similarly, Lazareva and Wasserman found that participants who were able to discriminate the position of a stimulus within a given series (through the provision of ordered feedback) were significantly more likely to demonstrate TI. The conclusions from both studies indicate that enhancing awareness of the structure of the hierarchy facilitates the emergence of TI. Such an intervention, however, would be counterproductive for the purposes of the present study. Recall that the ToF effect is said to take place when the relation between two or more stimuli has been inferred, or 'derived' (Hayes et al., 2001), and an awareness of the hierarchical or ordered relationship is part of what becomes derived through the ToF process. Therefore, any additional procedure that would have enabled the ordering of the series could have confounded the present findings by obscuring the emergence of ToF (and, presumably, awareness of the relational hierarchy) from conditional discrimination training alone (cf. Barnes & Roche, 1996).

A final issue worth considering is the inconsistent ratings observed for stimulus C among some of the participants who demonstrated TI (see P8 and P9 in Fig. 4). Specifically, P8

---

[4] According to the expression $(n\ k)\ P^k\ q^{n-k}$ for binomial probability, where k is the number of predictions met, P is the chance probability of meeting a prediction ($P = .5$), n is the number of predictions made, and q is the probability of failure in one trial ($1 − p$).
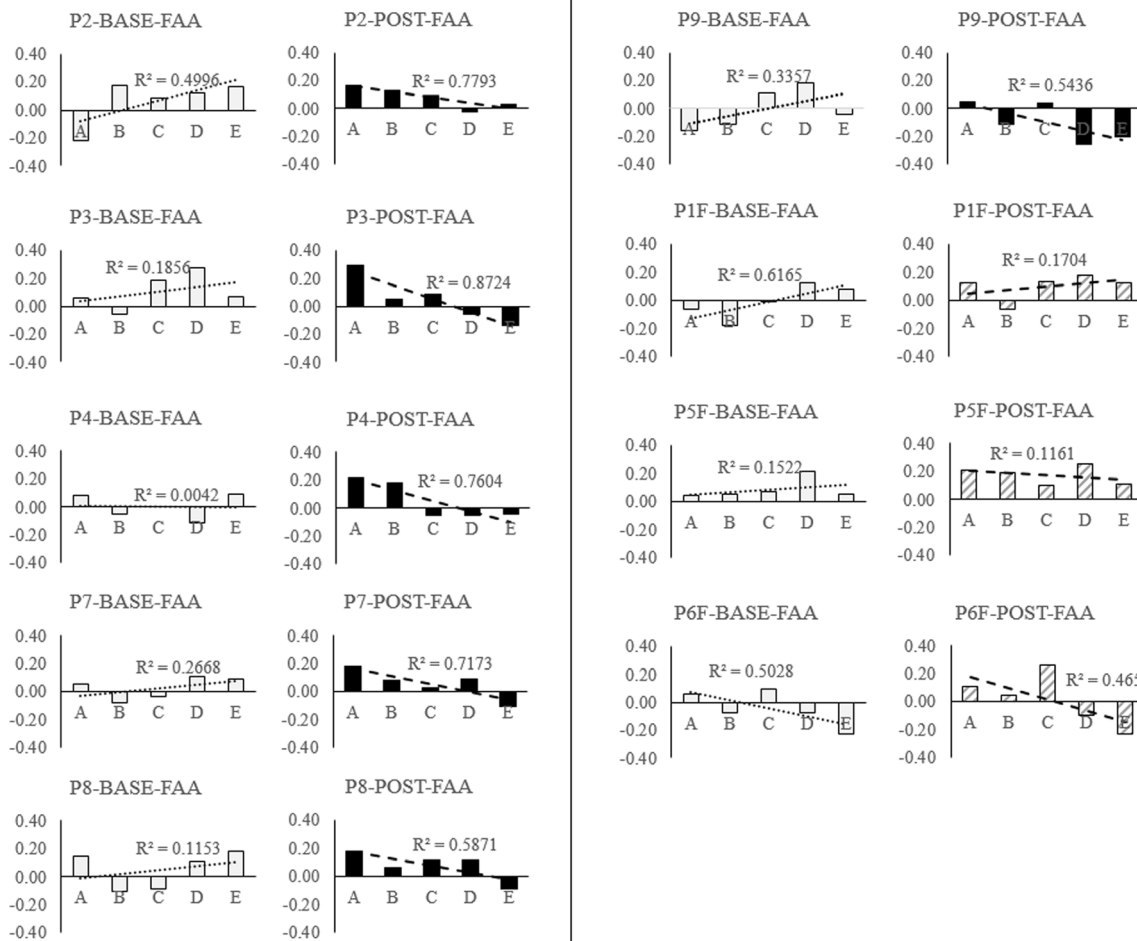
Fig. 6 Frontal alpha asymmetries (FAAs) per individual participants. Positive values indicate greater left prefrontal activation; negative values indicate greater right prefrontal activation.

rated C as unhappier than D, whereas P9 rated C as happier than D. Although this may indeed be an artifact of a small sample and low numbers of training and testing trials, the issue does not detract from the present findings for at least two reasons. First, because C was the intermediate stimulus in the relational hierarchy, no ToF was specifically predicted for that stimulus. Indeed, Table 2 confirms that the pre-TI to post-TI changes for the C stimulus were the least incremental and the most variable. Second, all participants who demonstrated TI provided ratings that fell along a gradient, as would be predicted by an A >> E structure (see Fig. 4), indicating that ToF had taken place across the hierarchy as a whole. Future researchers could refine the present protocol in order to assess whether stimulus functions derived for members of an ordered series would be hierarchically ordered as well.

The present study indicates that specific instances of contextually controlled relational responding (CCRR) performance in humans may not only yield function transformation, but potentially function *induction* as well (Baum, 2012). That is, when topographically relevant stimuli (e.g., masked faces; Amd & Roche, 2015, p. 539) are related in the presence of

contexts that specify a comparative, emotional dimension (e.g., a *happier-than* cue), the emotional functions of a stimulus may be enhanced with relation not to only another stimulus, but also in regards to its previous state (see Fig. 4). If one accepts permits the assumption that emotional ToF processes are to a large part driven via derivation in precisely specified contexts, some potential implications for verbal behavior are worth considering. Consider, for instance, the role of "rumination" in depressed individuals. In many cases, depressed individuals are unable to forgo emotionally aversive elements of their verbal repertoire, even when their surrounding contexts are no longer directly averse (Hetherington & Moulds, 2013). If the present speculation holds, one reason for the maintenance of ruminating behaviors may be that afflicted individuals become "entangled" in patterns of CCRR, in which the specific structure of transitively related stimuli/events can serve to maintain, transform, and perhaps induce emotional functions across said events almost indefinitely. Future researchers could explore this hypothesis in a number of ways: First, since in the present experiment we trained a five-member series, other researchers could attempt

to expand the number of members in the series to see whether the transformation effects presently reported would hold (but see Amd & Roche, 2015). Second, researchers could try embedding a "meaningful" emotion-eliciting stimulus at different locations within the established relational series (cf. Doran & Fields, 2012; Nedelcu, Fields, & Arntzen, 2015). Doing so would allow for the assessment of ToF gradients across comparatively related stimuli in accordance with various methodological parameters, such as nodal distance (Fields, Adams, Verhave, & Newman, 1990). Third, researchers could try "merging" different types of relational structures to assess whether functions transformed in the predicted directions. For instance, after establishing the equivalence class A–B–C and the comparative hierarchies D > B > G and X < B < Y (where B is the stimulus common to all three functional classes), one could assess whether the functions of D would be more than A, and so forth (see Mackay, Wilkinson, Farrell, & Serna, 2011, for a seminal study on class mergers). With regard to clinical work, researchers could establish stimulus–stimulus relations in the presence of *happier-than* contexts across ruminating versus nonruminating individuals in order to explore for differences in the rates of relation acquisition, response latencies/accuracies, and ToF following the demonstration of TI. Such findings could illuminate the extent to which deriving stimulus relations in valence-setting contexts contributes to pathological verbal behavior.

References

Allen, J., Coan, J., & Nazarian, M. (2004). Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biological Psychology, 67,* 183–218.

Amd, M., & Barnes-Holmes, D. (2014). A derived transformation of evaluative preferences using implicit association tests. *Psychological Record, 64,* 475–485. doi:10.1007/s40732-014-0033-z

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,* 318–334. doi:10.1002/jeab.19

Amd, M., & Roche, B. (2015). A derived transformation of valence functions across two 8-member comparative relational networks. *Psychological Record, 65,* 523–540. doi:10.1007/s40732-015-0128-1

Barnes, D., & Roche, B. (1996). Stimulus equivalence and relational frame theory are fundamentally different: A reply to Saunders' commentary. *Psychological Record, 46,* 489–508.

Baum, W. M. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior, 97,* 101–124. doi:10.1901/jeab.2012.97-101

Bazanova, O. M., & Vernon, D. (2014). Interpreting EEG alpha activity. *Neuroscience & Biobehavioral Reviews, 44,* 94–110. doi:10.1016/j.neubiorev.2013.05.007

Boyle, S., Roche, B., Dymond, S., & Hermans, D. (2015). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion,* 1–3. doi:10.1080/02699931.2014.1000831

Briesemeister, B., Tamm, S., Heine, A., & Jacobs, A. (2013). Approach the good, withdraw from the bad—A review on frontal alpha asymmetry measures in applied psychological research. *Psychology, 4,* 261–267. doi:10.4236/psych.2013.43A039

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76,* 387–404. doi:10.1037/h0027578

Coan, J. A., & Allen, J. J. B. (2003). Frontal EEG asymmetry and behavioral activation and inhibition systems. *Psychophysiology, 40,* 106–114.

Coan, J. A., & Allen, J. J. B. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology, 67,* 7–49. doi:10.1016/j.biopsycho.2004.03.002

Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: Perspectives on frontal EEG asymmetry research. *Biological Psychology, 67,* 219–234. doi:10.1016/j.biopsycho.2004.03.008

Delius, J., & Siemann, M. (1998). Transitive responding in animals and humans: Exaptation rather than adaptation? *Behavioural Processes, 42,* 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,* 243–256. doi:10.1901/jeab.2012.98-243

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,* 179–197. doi:10.1901/jeab2007.45-05

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst/ MABA, 23,* 239–254.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53,* 345–358. doi:10.1901/jeab.1990.53-345

Fox, N. A., & Davidson, R. J. (1986). Taste-elicited changes in facial signs of emotion and the asymmetry of brain electrical activity in human newborns. *Neuropsychologia, 24,* 417–422. doi:10.1016/0028-3932(86)90028-X

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. doi:10.1037/0022-3514.85.2.197

Hagemann, D. (2004). Individual differences in anterior EEG asymmetry: Methodological problems and solutions. *Biological Psychology, 67,* 157–182. doi:10.1016/j.biopsycho.2004.03.006

Hagemann, D., Naumann, E., & Thayer, J. F. (2001). The quest for the EEG reference revisited: A glance from brain asymmetry research. *Psychophysiology, 38,* 847–857.

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological Psychology, 84,* 451–462. doi:10.1016/j.biopsycho.2009.08.010

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 Press.

Hetherington, K., & Moulds, M. L. (2013). Does mode of processing during a positive experience have consequences for affect?

Journal of Behavior Therapy and Experimental Psychiatry, 44, 165–171. doi:10.1016/j.jbtep.2012.10.002

Hinojosa, J. A., Carretié, L., Valcárcel, M. A., Méndez-Bértolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective, & Behavioral Neuroscience, 9,* 173–189. doi:10.3758/CABN.9.2.173

Huang, Y., Zhou, R., Cui, H., Wu, M., Wang, Q., Zhao, Y., & Liu, Y. (2014). Variations in resting frontal alpha asymmetry between high-and low-neuroticism females across the menstrual cycle. *Psychophysiology, 52,* 182–191. doi:10.1111/psyp.12301

Ille, N., Patrick, B., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology, 19,* 113–124. doi:10.1097/00004691-200203000-00002

Kawahara, S., & Shinohara, K. (2012). A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Building bridges across cognitive sciences around the world: Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 569–574). Austin, TX: Cognitive Science Society.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29,* 169–195. doi:10.1016/S0165-0173(98)00056-3

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. *Animal Cognition and Emotion, 24,* 1377–1388. doi:10.1080/02699930903485076

Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task awareness on human performance. *Behavioural Processes, 83,* 99–112. doi:10.1016/j.beproc.2009.11.002

Light, G. A., Williams, L. E., Minow, F., Sprock, J., Rissling, A., Sharp, R., . . . Braff, D. L. (2010). Electroencephalography (EEG) and event-related potentials (ERPs) with human participants. In J. N. Crawley et al. (Ed.), *Current protocols in neuroscience* (Unit 6.25, pp. 1–24). New York, NY: Wiley. doi:10.1002/0471142301.ns0625s52

Mackay, H. A., Wilkinson, K. M., Farrell, C., & Serna, R. W. (2011). Evaluating merger and intersection of equivalence classes with one member in common. *Journal of the Experimental Analysis of Behavior, 96,* 87–105. doi:10.1901/jeab.2011.96-87

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

Nedelcu, R., Fields, L., & Arntzen, E. (2015). Arbitrary conditional discriminative functions of meaningful stimuli and enhanced equivalence class formation. *Journal of the Experimental Analysis of Behavior, 103,* 349–360.

O'Hora, D., Barnes-Holmes, D., & Stewart, I. (2014). Antecedent and consequential control of derived instruction following. *Journal of the Experimental Analysis of Behavior, 102,* 66–85.

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,* 263–283.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the Event-Related Potential. *International Journal of Medical Sciences, 2,* 147–154.

Scherg, M., & Berg, P. (1991). Use of prior knowledge in brain electromagnetic source analysis. *Brain Topography, 4,* 143–150.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide, version 2.0.* Pittsburgh, PA: Psychology Software Tools.

Steirn, J. N., Weaver, J. E., & Zentall, T. R. (1995). Transitive inference in pigeons: Simplified procedures and a test of value transfer theory. *Animal Learning & Behavior, 23,* 76–82. doi:10.3758/BF03198018

Sutton, S. K., & Davidson, R. J. (1997). Prefrontal brain asymmetry: A biological substrate of the behavioral approach and inhibition systems. *Psychological Science, 8,* 204–210. doi:10.1111/j.1467-9280.1997.tb00413.x

Tonneau, F. (2001). Equivalence relations: A critical analysis. *European Journal of Behavior Analysis, 2,* 1–33. doi:10.1080/15021149.2001.11434165

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

Von Fersen, L., Wynne, C. D., Delius, J. D., & Staddon, J. E. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 334–341. doi:10.1037/0097-7403.17.3.334

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

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning & Behavior, 23,* 207–217. doi:10.3758/BF03199936