An Examination of Transformation of Evaluative and Consequential Functions through Derived Relations with Participant-Generated Values-Relevant Stimuli

Emily K. Sandoz[1]

Michael J. Bordieri[2]

Gina Q. Boullion[1,3]

Ian Tyndall[4]

[1](Corresponding author) University of Louisiana at Lafayette, P.O. Box 43644, Lafayette, LA 70504; emily.sandoz@louisiana.edu

[1,3]University of Louisiana at Lafayette

[2]Murray State University

[3]Author has moved to University of Mississippi since the work described was completed.

[4]University of Chichester

Acknowledgements: This research was funded, in part, by a Research Competitiveness Subprogram Grant LEQSF(2011-14)-RD-A-29 through the Louisiana Board of Regents. The authors would also like to thank Emmie R. Hebert and Shelly Greene for their assistance with data collection, and Nolan Williams for providing feedback with the literature review for a previous version of this manuscript.

Abstract

Values-affirmation interventions have demonstrated efficacy in increasing approach behavior in the context of potential threat. In other words, writing about values seems associated with changes to the functions of previously aversive events. Evaluative conditioning and derived relational responding have been offered as possible mechanisms by which values interventions change behavior. The current study aimed to extend the extant literature by demonstrating derived relational responding and subsequent transformation of evaluative and consequential functions with values-relevant stimuli. Participants were 34 undergraduate students. Participants generated personally meaningful values-relevant stimuli after engaging in a values-affirmation task and were subsequently trained through matching to sample to coordinate a subset of those stimuli to arbitrary stimuli. All participants exhibited mutual entailment, and all but one exhibited combinatorial entailment, suggesting that individuals learn to coordinate events with values quite readily. Further, there was evidence of transformation of functions, both in terms of changes in ratings of derived stimuli and in terms of changes in approach and escape behavior. These data are offered in support of continued scientific exploration of what values are, how they emerge, and how they are best intervened upon.

Keywords: values; derived relational responding; Relational Frame Theory; verbal behavior; transformation of function

Highlights

- Participants related self-generated values words to arbitrary stimuli across derived relations.
- Participants demonstrated derived transformation of evaluative functions.
- Participants demonstrated derived transformation of consequential functions.
- Entailment and transformation of function can be modeled with participant-specific stimuli.

An Examination of Transformation of Evaluative and Consequential Functions through Derived Relations with Participant-Generated Values-Relevant Stimuli

Villatte (2020) defined values as overarching and intrinsic sources of positive reinforcement. Based in a Relational Frame Theory (RFT; Hayes et al., 2001) perspective, this definition proposes values as positive in that they are appetitive rather than aversive, intrinsic in that reinforcement is inherent in the valued behavior because of its verbal (i.e., symbolic) relation with a specific value, and overarching in that topographically different actions might be reinforced via connection to that particular value. Despite the appeal of this definition, little basic empirical work has been conducted within an RFT framework as to how values-based action emerges in a person's learning history or how values-based stimuli influence action. Nonetheless, an RFT-based account has the potential to clarify underlying behavioral processes by which stimuli come to function as 'values' that might guide appetitive or approach behavior, such as that seen in social psychology values-affirmation literature (e.g., Cohen & Sherman, 2014).

Values (or self-) *affirmation* generally refers to the impact of brief values writing or reflection interventions focused on increasing approach behavior in the context of potential threat (e.g., risks to health, safety, or self-evaluation; Cohen & Sherman, 2014). For example, sexually active people given the opportunity to write about an important value are more likely to purchase condoms after an AIDS educational video than those who write about an unimportant value (Sherman et al., 2000). College women who write about important values significantly outperform in a college physics class those who do not, with differences being most pronounced for those who endorse stereotypes regarding women underperforming men in physics (Miyake et al., 2010). In another example, people who write about highly rated values are more likely than those who write about low rated values to help others succeed in ways that are personally threatening (Tesser et al., 1996). Values-affirmation has been observed even when beliefs suggest a lack of openness. For example, climate change skeptics who are first given the opportunity to write about important values respond to a message on anthropogenic climate change by describing themselves as more able to act to prevent it (Prooijen & Sparks, 2014).

As an example, Peters et al. (2017) highlights the typical values-affirmation procedure that has been successfully applied across diverse domains (Cohen & Sherman, 2014). It also represents the perspective that engaging in the values-affirmation procedure functions to protect the self and identity, which generalizes well beyond the affirmation task itself to key contexts (see Cooke et al., 2012). Peters et al. (2017) examined whether a values-affirmation task administered to students on a statistics module at the beginning of the semester would positively impact numeracy ability, self-perceptions, and attitudes. Indeed, Peters et al. (2017, e0180674) predicted that engaging in the values-affirmation exercise would help "stave off a recursive cycle of experienced threat from the course and improve development of objective numeracy skills", stating "we also expected improvements to…protection of self-perceptions about ability and attitudes towards numeric information." The mechanism by which this expected effect might occur was not clearly specified. Peters et al. stated that reflecting on core values can help people (1) focus on their longer-term goals in life and deflect from pressing current concerns and pressures and (2) accept thoughts which are counter to their attitudes towards the behavior of interest (e.g., health behaviors; school work). Peters and colleagues provided participants with the standard values-affirmation task instructions as they were asked to rank a list of six values (art/music/theatre, science/pursuit of knowledge, relationships with family/friends, government/politics, spiritual/religious values, business economics) by personal importance. The experimental group (n = 112) were told to write about why their most important value was meaningful and important to them (i.e., values-affirmation), while the control group (n = 109) were asked to write about how their least important value might be meaningful and important to other people. Both groups then selected the top two reasons why their chosen value was important to them (values-affirmation) or to others (control). Thus, the task was self-relevant only for the values-affirmation group. The results were somewhat mixed over a range of dependent measures, but Peters et al. (2017) concluded that this values-affirmation intervention (importantly, that was not statistics or numeracy related) produced "positive, albeit small, differences over time for subjective and objective numeracy and generalized to the seemingly unrelated domains of financial literacy and health-related behaviors" (e0180674).

In summary, the values-affirmation literature suggests that the functions of aversive events seem to change when people have just previously written about important personal values. It is unclear, however, *how* this change takes place. The present study aims to examine potential behavioral processes that might underpin such important symbolic change in functions of aversive or appetitive events, namely transformation of evaluative and consequential stimulus functions.

Evaluative Conditioning

Conditions under which the meaning or functions of an event shift have been more broadly considered in terms of evaluative conditioning (EC), where the valence of a stimulus changes due to the pairing of that stimulus with another stimulus (see De Houwer, 2007). Of particular relevance to values-affirmation may be the generalizability of EC absent direct pairing between stimuli (Amd & Roche, 2016). More specifically, generalization of EC has been reported via transformation of function through derived relational responding (e.g., Barnes-Holmes et al., 2000; Dack et al., 2010; Smyth et al., 2006; Valdivia-Salas et al., 2013). Derived relational responding (DRR) is described in RFT as the process by which humans respond to stimuli based on their arbitrarily applied relations to other stimuli. Verbally-able humans are able to relate stimuli based in part on arbitrarily applicable contextual cues that determine what sort of relating is likely to be reinforced (e.g., responding to a U.S. dime as if it is “bigger than” a nickel in the context of how much candy can be purchased despite being smaller in size). DRR, then, is offered as a behavior analytic account of symbolic behavior with implications for how it is that stimuli come to control behavior despite little or no direct learning history with such stimuli.

From an RFT perspective, DRR is a generalized operant, meaning that with repeated reinforcement of this sort of relating across multiple exemplars, DRR emerges as a class of behaviors that are functionally similar but lack topographical similarity (Barnes-Holmes & Barnes-Holmes, 2000). Mutual entailment is the first property of relational framing: If we directly learn an F-G relation, we can derive the symmetrical G-F relation. For example, if we learn that F is “more than” G then we can derive that G is “less than” F . Combinatorial entailment is the second property: if we know an X-Y and a Y-Z relation, we can derive the respective mutually entailed relations (Y-X and Z-Y), but also the X-Z and Z-X relations (McLoughlin et al., 2020). DRR research has a robust evidence-based literature of laboratory-generated derived symbolic responding across numerous relations including, for example: comparison, opposition, and hierarchy (see Dymond & Roche, 2013; McLoughlin et al., 2020 for an overview).

The third property of relational framing is the transformation of stimulus function, which involves the alteration of functions of stimuli consistent with emergent relations (e.g., same as; opposite) within the derived relational network (e.g., Amd & Roche, 2015, 2016; Dymond et al., 2019; Dymond & Rehfeldt, 2000; Perez et al., 2017). Of relevance to the present study, transformation of evaluative functions was first demonstrated by Barnes-Holmes and colleagues (2000). Participants first learned to relate nonsense syllables, VEK and ZID, to CANCER and HOLIDAY, then to BRAND X and BRAND Y, respectively. Subsequently, participants rated cola labeled BRAND Y more favorably than identical cola labeled BRAND X. Similar findings have been demonstrated with a range of stimuli (see Hofmann et al., 2010), including negatively-valenced evaluations. For example, participants reported fear and disgust toward a nonsense syllable after having related that nonsense syllable to another nonsense syllable that had been paired with images of spider attacks (Smyth et al., 2006).

Within this framework, empirical investigations of transformation of consequential functions are also relevant, as approach or selection behavior represents a more direct measure of stimulus valence than participant ratings or reports. Reinforcing and punishing functions, once directly conditioned to one member of a relational class, have been demonstrated across all members of the class (Hayes et al., 1991). Similarly, reinforcing and punishing functions have been transformed across Same/Opposite (Whelan & Barnes-Holmes, 2004) and More-than/Less-than (Whelan et al., 2006) relations. Valdivia-Salas and colleagues (2013) demonstrated transformation of consequential functions even with abbreviated testing, and without contingent presentation of derived consequences or interspersions of conditioning trials.

In this way, RFT may offer an account of how it is that the opportunity to write about important values could transform evaluative and consequential functions of events. RFT has been applied to the conceptualization of values as they are employed in the therapeutic context. One such analysis emphasizes a value as a primary node in a hierarchical relational network including lower levels of abstract consequences, long-term goals, and varied but specific behavioral patterns that may contribute to accomplishing those goals (Plumb et al., 2009). Transformation of function across such a network can be described in terms of *augmenting* (see Kissi et al., 2017), a form of rule governed behavior where the rule (e.g., a stated value) impacts the extent to which stimuli or events in a person's environment function as reinforcers or punishers. In this way, consequences intrinsic to valued behaviors can come to maintain them (Wilson et al., 2011). The same process described clinically might be relevant in values-affirmation procedures.

Augmenting can be contrasted with other forms of verbal control, where rules about what should be pursued or what must be avoided result in behavior that is rigid and insensitive to direct consequences (McCracken et al., 2014). For example, *pliance* is a form of rule-governed behavior that is under control of socially-mediated reinforcement for correspondence between the rule and relevant behavior rather than intrinsic consequences of the behavior (Kissi et al., 2017).

Rigidity and insensitivity due to a dominance of verbal functions has also been described as *fusion* (Assaz et al., 2018; Hayes, 2004), which has been associated with psychological difficulties such as anxiety, depression, and rumination (Gillanders et al., 2014). Rigid constructions about what should be pursued as a value or how a value must be pursued can limit effectiveness of, and sensitivity in, responding. This has been referred to as *values fusion* (Hayes et al., 2012; p. 318).

To our knowledge, there is no basic laboratory empirical research that examines how values-affirmation or values fusion might function as a process. The present study represents a tentative first attempt to explore this process from an RFT standpoint. Values writing has been shown to facilitate participant production of stimuli they then rate as meaningful, evocative, and reminiscent of something important (Sandoz & Hebert, 2015). When participants write about important values, this seems to create a context for (1) relating values to the stimuli generated (i.e., the words they write), and (2) relating those stimuli to other, values-relevant events, potentially allowing for a transformation of function of those events such that they are more likely to increase in saliency, be evaluated positively, and be approached more frequently. The present study employed a values writing task based on the most common values-affirmation procedure in order to generate participant-specific values stimuli (McQueen & Klein, 2006). It was hypothesized that participants would demonstrate (1) mutual and combinatorial entailment of relations among participant-specific stimuli and arbitrary stimuli, and (2) transformation of pre-experimental evaluative and consequential functions of arbitrary stimuli for consistency of emergent relations. If successful, this will be the first study to demonstrate these relational processes in the context of values, and using participant-specific, and empirically selected values stimuli.

Method

Participants

The sample was comprised of 34 undergraduate students recruited from a Southern University in the United States. Participants were 71% female with 71% self-identifying as White/Caucasian, 24% as African American, and 5% as Multiracial/Other. The mean age was 19.8 (*SD* = 2.3). The experimental protocol was approved by the first author’s Institutional Review Board (IRB) prior to participant contact and informed consent was obtained from all participants.

Apparatus and Setting

Nine Dell Optiplex 755 computers, outfitted with 2200.0 MHz Intel Core 2 Duo E4500 processors, were used along with their 15×12-inch monitors, keyboards, and mice. Instructions and stimuli were displayed on the monitor and all responses were recorded in terms of rate and accuracy. The computer task was designed using Visual Basic 2008. Participants completed the computer task in a 25’ by 30’ computer laboratory, isolated from noise and other distractions. Participants were seated at desks with privacy screens (30” wide, 15” tall, and 22” deep), and the computers used for the study were arranged such that every other desk was empty.

Procedure

Phase 1: Values Writing

Participants were provided with descriptions of common areas of life that people value, including theoretical, economic, aesthetic, social, political, and religious values (McQueen & Klein, 2006). They were then provided the following prompt: “Please write about your most deeply held values for ten minutes. You will have the choice of whether or not you want to share your writing with the experimenters, so be sure to write about values that are personally meaningful to you. When you are ready click Begin.” Participants typed their responses in a text box and were allowed to write uninterrupted for ten minutes. They were given no additional guidance regarding how many life domains to consider during the task. Upon completion of the values writing task participants were given the option of allowing their writing to be retained for sharing with the experimenter by selecting “yes” or “no.” Once the participants had finished writing about their values they pressed “OK” to continue to Phase 2.

Phase 2: Stimuli Generation

Upon beginning Phase 2, participants were instructed to provide nine words in three categories. First, participants were provided with the instructions, “Please select a word from what you have written that represents what you value. Write the word in the space below.” After providing an initial response, participants were asked to provide two additional value words for a total of three values words. Next participants were provided with the instructions, “Please think of a value that you do not find particularly meaningful but that you would feel guilty or ashamed about if others knew it was not very important to you.” Participants provided three words in this category that represented values that might be endorsed because of pliance (i.e., fused value words). Finally, participants were provided with the instructions, “Please think of a value that you do not find particularly meaningful and do not care if others knew it was not very important to you” and provided three more words that represented non-values (i.e., neutral, but values-relevant words). For each instruction set, participants read the instructions, typed the three words into blank text boxes, and clicked “Continue” to move on to the next selection. For each word entry, the program rejected single letter responses, duplicate responses, multiple words in a response (i.e., the presence of a space), and non-alphabetic characters (i.e., numbers or symbols). Rejected responses resulted in the participant being re-prompted to enter a valid word. Otherwise, the program accepted all idiographic responses regardless of their content. Once the participants had provided all nine words, they clicked “Continue” to continue to Phase 3.

Phase 3: Stimulus Function Pretesting

During this phase, participants rated the “meaningfulness” and “difficulty” of six arbitrary shapes (potential F stimuli) along with the nine idiographic words and three experiment provided words (“machine,” “pencil,” and “address;” the 12 words collectively potential E stimuli) on a visual analog scale ranging from 0 to 100. Each of the 18 stimuli were presented on the screen one at a time and participants were provided with the prompt, “How meaningful is this to you?” along with the visual analog scale with the anchors of “Not at all meaningful” and “Very meaningful.” All stimuli were then presented once again along with the prompt, “How much difficulty does this cause for you?” along with the visual analog scale with the anchors of “No difficulty” and “Extreme difficulty.” For each stimulus, participants were to drag the pointer across the visual analog scale to reflect the amount of meaningfulness or difficulty elicited from that stimulus (scale ranged from 0 to 100). Words included the three value words, three fused value words, and three neutral value words provided in Phase 2 along with three experimenter generated words (i.e., “machine,” “pencil,” and “address”). These three words were arbitrarily generated by the researchers as potential neutral words with regard to values and valuing. They were included in the design to increase the likelihood of having stimuli rated low in meaningfulness and difficulty to select from in the subsequent stimulus selection procedure.

Stimulus Selection

Following the initial stimuli ratings, a stimuli selection algorithm programmed by the experimenters generated a unique set of E1, E2, E3, F1, F2, and F3 stimuli for each participant’s matching to sample task. The E1 (Value) stimulus was chosen by reviewing the participant’s meaningfulness ratings of each the 12 words and then selecting the word with the highest meaningfulness rating. In the case of a tie (e.g., multiple words rated as 100), the algorithm selected the word ranked most recently by the participant. While difficulty and vulnerability are conceptually linked to valuing (Sandoz & Anderson, 2015), the algorithm considered only meaningfulness in selecting a value stimulus to ensure that the selected stimuli were highly valanced with regard to meaningfulness. The E2 (Fused Value) stimulus was chosen by calculating a discrepancy score (difficulty – meaningfulness) for each of the 12 words and then selecting the word with the highest discrepancy score. The E3 (Neutral) stimulus was chosen by calculating an overall meaningfulness and difficulty score (meaningfulness + difficulty) for each of the 12 words and then selecting the word with the lowest overall score. This was done to select the textual stimuli with the least combined meaningfulness and difficulty valance for each participant (i.e., the most neutral score). Likewise, the F stimuli were chosen by calculating an overall meaningfulness and difficulty score (meaningfulness + difficulty) for each of six possible F stimuli and then selecting the three with the lowest overall score. The lowest scoring F stimulus was selected as F1, the penultimate as F2, and the third lowest as F3.

Phase 4: Matching to Sample

Following completion of stimulus ratings, participants engaged in a computer task training relational responding using a one-to-many matching-to-sample conditional discrimination task with values-relevant stimuli and arbitrary shapes (see Figure 1). Stimuli included three three-member classes (D, E, & F). The D stimuli were arbitrary shapes that were consistent across participants. The E and F stimuli were selected according to the procedure described above.

Conditional discrimination training consisted of a stimulus at the top of the screen (D1, for example) and three comparison stimuli across the bottom of the screen (E1, E2, and E3, for example). Participants were instructed to select a stimulus from the bottom array by clicking. During training, selection was followed by the presentation of the words, “correct” or “incorrect” on the screen for 1.5 seconds. During testing, selection was followed by a blank screen for 1.5 seconds.

There were five phases in the conditional discrimination procedure including three phases training relational responding and two phases testing for mutual and combinatorial entailment. The procedure employed a one-to-many procedure. Specifically, the first stage trained D-E relations of coordination (D1-E1, D2-E2, D3-E3), and the second stage trained D-F relations of coordination (D1-F1, D2-F2, D3-F3). For both stages, participants had to correctly complete 16 out of 18 trials (89%) to move on to the next stage. The third stage was a mixed training including both D-E and D-F relations. In this stage 32 of 36 trials (89%) had to be completed correctly for participants to move on to testing.

Next, relational testing probed for mutual entailment of derived E-D and F-D relations, and combinatorial entailment of E-F/F-E relations. Testing criterion was 16/18 trials (89%) for both stages. If participants did not meet criterion for combinatorial entailment, they returned to the mixed D-E/D-F training stage. If participants did not meet the criterion for combinatorial entailment a second time, they were dismissed from the study.

Phase 5: Stimulus Function Post Testing

Following the conditional discrimination task, participants again rated the meaningfulness and difficulty of the three E and three F stimuli with a procedure identical to that in Phase 3.

Phase 6: Approach and Escape Task

In addition to rating stimuli, participants performed a task designed to approximate approach and escape responses with all 9 study stimuli (i.e., D1, D2, D3, E1, E2, E3, F1, F2, and F3[1]). Approach and escape have been associated with pulling toward and pushing away, respectively (Chen & Bargh, 1999). The current study adapted the computerized Approach Avoidance Task (Rinck & Becker, 2007) to allow for more simple quantification of approach and escape behavior, replacing the use of a joystick with a typical keyboard and instructions to use 'F' and 'J' keys to "pull toward" or "push away." Specifically, participants were provided the instructions, "During the next phase of the study one image will be presented on the screen at a

[1] D, E, and F were used to denote classes instead of the conventional A, B, and C as this study was part of a larger series of studies that did not repeat class labels across sets of stimuli. We retain these labels here not only for consistency between our data and the record, but also for consistency among studies in this series.

time. After viewing the image for a few seconds, you will have the ability to modify the image. To pull the image closer to you press the ‘J’ key. To push the image away from you press the ‘F’ key. If you do not wish to change the image you can simply not press any key.” Stimuli were presented one at a time with the instructions “Press ‘F’ to make smaller, Press ‘J’ to make bigger.” Each trial began with a 2-second display where responding was not possible followed by a variable 5 to 10-second window where responding was possible. “Approaching” stimuli involved pressing the J button on the keyboard, which was consequated by an increase in the stimulus size from 300 pixels by 30 pixels per response. “Escaping” stimuli involved pressing the F button, which was consequated by a reduction in size, from 300 pixels, by 30 pixels per response. Participants could approach each stimulus by increasing the size to a maximum size of 600 pixels or could escape each stimulus by decreasing the size until it no longer remained on the screen. Additionally, once a response was selected (i.e. either escape or approach) the other response option was no longer available, so that participants could only exhibit escape or approach responses in each trial. At the beginning of this phase participants were given four practice trials with corrective feedback (i.e., “Correct!” or “Please follow the instructions on the screen.”). During two of the practice trials participants were instructed to “pull closer” and during the other two trials they were instructed to “push away.” Following these practice trials participants were exposed to each of the nine study stimuli three times in a random order. At the end of each trial (i.e., the 5-10 second variable window), if the stimulus had not been removed by the participant, it was removed for a 500 millisecond intertrial interval. Following the ITI, an orienting response (“Press Spacebar to Continue”) was required to start the next trial, which commenced after a 1000 second pre-trial interval.

Analytic Strategy

All study analyses were conducted using SPSS version 21 and R version 3.5.0. Data were screened for missing and out of range values prior to analysis.

Stimuli Generation and Initial Stimuli Ratings

Descriptive statistics were calculated for the values writing and stimuli generation phases of the study to allow for the evaluation of the participant-generated stimuli. In particular, word count of the writing task and the frequency of commonly identified values stimuli were explored.

Stimuli Selection

The performance of the stimuli selection algorithm was evaluated by assessing the quality of the stimuli selected for use in the main experimental task. In particular, descriptive statistics of meaningfulness ratings for the Value stimulus (E1), discrepancy ratings (difficulty - meaningfulness ratings) for the Fused Value stimulus (E2), and an overall rating (meaningfulness + difficulty) ratings for the Neutral stimulus (E3) were assessed. In addition, the generation sources (i.e., values words, fused value words, neutral values words, or experimenter generated words) of the assigned stimuli for each class member were explored.

Class Acquisition

Descriptive statistics of class acquisition performance on the matching to sample task were calculated for percent accurate responding during testing phases and trials blocks to criterion during training phases. Trial blocks to criterion were calculated for each participant by summing the number of times they were sequenced through the training block before meeting the pass criterion (≥ 89%). Descriptive statistics of training time as well as frequency counts of overall pass/fail status were also analyzed.

Meaningfulness and Difficulty Ratings

Differences in meaningfulness and difficulty ratings of combinatorially entailed stimuli before and after class acquisition training were assessed via two repeated measure MANOVAs. For the *meaningfulness* model, pre-meaningfulness ratings were entered as Time 1 scores and post-meaningfulness ratings were entered as Time 2 scores for the derived Value (F1), derived Fused Value (F2), and derived Neutral (F3), stimuli, respectively. The *difficulty* model was identical with the exception that difficulty ratings were entered. Significant repeated measure MANOVA models were followed up by a series of univariate repeated measure ANOVAs across each of the stimulus classes (i.e., F1, F2, and F3). In addition, visual analysis of participant level data was conducted for combinatorially entailed meaningfulness and difficulty ratings.

A series of one-sample equivalence analyses were conducted using TOSTER (Lakens, 2018) to evaluate the equivalence of meaningfulness and difficulty ratings across stimuli post-class acquisition. A medium effect size (Cohn's *d* = 0.50) was chosen as the smallest effect size of interest (SESOI) as the sample size needed to achieve 80% power within the equivalence bounds (*N* = 35) closely matched the obtained sample size (*N* = 34).

Approach and Escape Responding

Approach and escape responses across the three trials for each study stimulus were combined into composite scores, with all approach responses to a stimulus scored positively and all escape responses to a stimulus scored negatively. In particular, a composite score was generated for each participant for each of the nine study stimuli (i.e., D1-D3, E1-E3, and F1-F3).

These composite scores were used as a data reduction strategy as each participants' approach and escape responses to each stimulus was considered along a continuum of -30 to 30, with a positive score indicating a pattern of approach responses, a negative score indicating a pattern of escape responses, and a score of zero indicating a pattern undifferentiated or null responding.

A repeated measure MANOVA model across study stimuli was conducted to determine whether composite approach/escape scores differed by class member (i.e. Value, Fused Value, and Neutral). Significant multivariate findings were further explored by a series of three repeated measure ANOVAs across each level of derivation (i.e., direct [E class], mutually entailed [D class], and combinatorially entailed [F class]). Each of these follow-up univariate ANOVAs was accompanied by an orthogonal Helmert contrast analysis. This contrast analysis first compared composite approach/escape responding towards the Value stimulus to the mean of Fused Value and Neutral Stimuli responding and then compared composite responding towards the Fused Value stimulus to composite responding towards the Neutral Stimulus.

A series of one-sample equivalence analyses were conducted using TOSTER (Lakens, 2018) to evaluate the equivalence of composite approach/escape responding scores across study stimuli. A medium to medium-large effect size (Cohn's *d* = 0.55) was chosen as the smallest effect size of interest (SESOI) as the sample size needed to achieve 80% power within the equivalence bounds (*N* = 29) matched the obtained sample size for this analysis. Visual analysis of participant level data was also conducted to explore each participant's pattern of approach/escape responding across study stimuli.

Results

Stimuli Generation and Initial Stimuli Ratings

Thirty participants (88%) agreed to share their valued writing with the experimenters. Among these participants, the average length of their 10-minute writing samples was 208.0 words (*SD* = 83.5) with a range of 21 to 376 words. With regard to the three value words generated by each participant, the most common words included “family” (14.7% of words generated), “religion” (5.9%), “education” (5.9%), “life” (4.9%), and “relationships” (3.9%). The most common fused value words included “religion” (8.8%), “environment” (6.9%), “work” (4.9%), “politics” (3.9%), and “money” (3.9%). The most common neutral words included “politics” (7.8%), “spirituality” (2.9%), “parenting” (2.9%), and “art” (2.9%).

Stimuli Selection

For the E1 (Value) stimulus, the average meaningfulness rating of the stimuli selected by the algorithm was very high (*M* = 97.6, *SD* = 5.5), indicating that the selected stimuli were high in meaningfulness functions. For 32 of the 34 participants (94%) the E1 stimulus was selected from one of their three values words, with 26 participants (76%) assigned the first values word they provided. Two participants (participants 11 and 13) rated an experimenter generated word as 100 on the meaningfulness VAS scale and were assigned the E1 stimulus of “machine” and “address,” respectively. For the E2 (Fused Value) stimulus, the average discrepancy score (difficulty – meaningfulness) of the stimuli selected by the algorithm was 49.3 (*SD* = 29.2), indicating that the selected stimuli were higher in difficulty functions relative to meaningfulness functions. Half of participants were assigned an E2 stimulus from one of their three generated fused value words while the other half were assigned an E2 stimulus from either one of their neutral (44%) or experimenter-generated words (6%). For the E3 (Neutral) stimulus, the average overall rating score (meaningfulness + difficulty) was low (*M* = 18.26, *SD* = 17.59), indicating that the selected stimuli were low in combined meaningfulness and difficulty functions. Fifty-three percent of participants were assigned an E3 stimulus from one of the experimenter-generated words, 26% from one of the neutral words, and 21% from one of the fused value words.

Class Acquisition

Participant performance during class acquisition is presented in Table 1. All participants earned a passing score on the test of mutual entailment and 31 participants (91%) achieved a passing score on the test of combinatorial entailment after initial class acquisition training. Of the three who did not pass, two achieved a passing score after exposure to a remedial block of mixed D-E/D-F training while one earned a score of 72% of the second test of combinatorial entailment. Data from this participant was removed from all subsequent study analyses.

Transformation of Stimulus Functions

Meaningfulness and Difficulty Ratings

Changes in meaningfulness ratings for the combinatorially derived stimuli (F Class) are presented in Figure 2. There was a significant difference in meaningfulness ratings across the combinatorially derived stimuli from pre to post class acquisition, $V(3, 30) =.901$, $p <.001$. Follow-up analyses revealed an increase in the meaningfulness of the F1 (Value) stimulus from pre-test (*M* = 14.30, *SD* = 14.53) to post-test (*M* = 85.03, *SD* = 26.44), $F(1, 32) = 238.90$, $p <.001$, partial $\eta^2 = .88$. No significant changes were observed for the F2 (Fused Value) or F3 (Neutral) stimuli. Participant level meaningfulness ratings of combinatorially derived stimuli (F Class) are presented in Figure 3. Visual analysis revealed that the majority of participants (*n* = 29; 88%)

displayed a substantial increase in meaningfulness ratings of the combinatorially derived values stimulus (F1). No clear pattern of ratings changes was noted for the meaningfulness ratings of the fused value (F2) or neutral (F3) stimuli.

Changes in difficulty ratings for the combinatorially derived stimuli (F Class) are presented in Figure 4. There was a significant difference in difficulty ratings across the combinatorially derived stimuli from pre to post class acquisition, $V(3, 30) = .5991$, $p < .001$. Follow-up analyses revealed an increase in the difficulty of the F2 (Fused Value) stimulus from pre-test ($M = 22.85$, $SD = 27.68$) to post-test ($M = 52.73$, $SD = 30.82$), $F(1, 32) = 26.63$, $p < .001$, partial $\eta^2 = .45$. No significant changes were observed for the F1 (Value) or F3 (Neutral) stimuli. Participant level difficulty ratings of combinatorially derived stimuli (F Class) are presented in Figure 5. Visual analysis revealed that the majority of participants ($n = 20$; 61%) displayed a clear increase in difficulty ratings of the combinatorially derived fused value stimulus (F1). No clear pattern of ratings changes was noted for the difficulty ratings of the value (F1) or neutral (F3) stimuli.

Descriptive statistics of post-class acquisition of meaningfulness and difficulty ratings of direct (E) and combinatorially derived (F) stimuli along with paired sample and equivalence t-test results are presented in Table 2. For the meaningfulness ratings of the values stimuli (E1-F1), the null-hypothesis test result was non-significant, and the equivalence test result was also non-significant. This pattern of findings indicates that the observed mean difference in meaningfulness was not statistically different from zero but also not statistically equivalent to zero within the bounds of a medium effect size, 90% CI [-0.33, 12.57]. For all other stimuli comparisons, the obtained findings were statistically not different from zero and statistically equivalent to zero within the bounds of a medium effect size.

Approach and Escape Responding

Participants emitted mostly correct responses (90.2%) during the four approach and escape practice trials with eight participants (24%) making one or more errors. Only four participants (12%) made multiple errors during the practice trials. They were removed from subsequent analyses with 29 participants retained for analysis.

A visual depiction of group level approach and escape response composites across all study stimuli is presented in Figure 6. There was a significant difference in approach and escape responding across class members (i.e., Value, Fused Value, and Neutral), $V(6, 23) = .918$, p <.001. Follow up analyses indicated that the differences in approach and escape responding across class members persisted across all levels of derivation: direct (E class), $F(2, 56) = 108.30$, $p < .001$, partial $\eta^2 = .80$; mutually entailed (D Class), $F(2, 56) = 64.73$, $p <.001$, partial $\eta^2 = .70$; and combinatorially entailed (F Class), $F(2, 56) = 66.03$, $p <.001$, partial $\eta^2 = .70$ stimuli. Follow-up Helmert contrast analyses across levels of derivation revealed that participants approached the Value stimulus (E1, $M = 25.79$, SD =7.92; D1, M =23.79, SD =11.42; F1, $M = 21.86$, SD =14.03) significantly more than the Fused Value stimulus and Neutral stimulus across all three levels of derivation: direct (E Class), $F(1, 28) = 243.74$, $p < .001$; mutually entailed (D Class), $F(1,28) = 143.35$, $p < .001$; and combinatorially entailed (F Class), $F(1, 28) = 119.47$, $p < .001$. Helmert contrasts revealed no significant differences in the degree to which participants escaped the Fused Value stimulus (E2, $M = -19.24$. $SD = 13.49$; D2, M =-14.86, SD =18.20; F2, M = -17.03, SD =15.23) and Neutral stimulus (E3, $M = -18.76$, $SD = 14.83$; D3, M =-18.76, SD

=13.98; F3, *M* =-18.17, *SD* = 14.47) across all three levels of derivation: direct (E Class), *F*(1, 28) = 0.02, *p* = .897; mutually entailed (D Class), *F*(1,28) = 0.81, *p* = .377; and combinatorially entailed (F Class), *F*(1, 28) = 0.09, *p* = .764.

Equivalence analyses of approach /escape composite scores across mutually and combinatorially entailed stimuli are presented in Table 3. Null-hypothesis test results (i.e., paired sample t-tests) across all stimuli pairings were non-significant, indicating that the obtained findings were statistically not different from zero. The mutually entailed relationship between the fused value word (E2) and D2 stimulus and the combinatorially entailed relationship between both the value word (E1) and F1 stimulus and fused value word (E2) and F2 stimulus were not statistically equivalent to zero within the bounds of a medium to medium-large effect size (*d* = .55). All other comparisons were statistically equivalent to zero within the bounds of a medium to medium-large effect size.

Participant level approach/escape responses across study stimuli are presented in Figure 7. Visual analysis revealed a clear and substantial pattern of approach responses to value stimuli across all levels of derivation for the majority of participants (*n* = 20; 69%). Four participants (14%) displayed a lower magnitude pattern of approach response (≤ 20 approach responses to each value stimulus), four participants (14%) displayed a pattern of approach/escape responses that differed markedly across levels of derivation, and one participant (3%) engaged in null responding. With regard to fused value stimuli, the majority of participants (*n* = 18; 62%) displayed a clear and substantial pattern of escape responses across all levels of derivation. Five participants (17%) displayed a pattern of approach/escape responses that differed markedly across levels of derivation, four participants (14%) engaged in a lower magnitude pattern of escape responses (≤ 20 approach responses to each fused value stimulus), and two participants (7%) engaged in a pattern of approach responses. With regard to the neutral stimuli, the majority of participants (*n* = 18; 62%) again displayed a clear and substantial pattern of escape responses across all levels of derivation. Four participants (14%) displayed a pattern of approach/escape responses that differed markedly across levels of derivation, four participants (14%) engaged in a lower magnitude pattern of escape responses (≤ 20 approach responses to each neutral stimulus), two participants (7%) engaged in a pattern of approach responses, and one participant (3%) engaged in null responding.

Discussion

The current study aimed to examine evaluative conditioning (EC) through derived relational responding (DRR) with values-relevant stimuli generated by the participant from a values writing task. Participants wrote about important values, selected key words from that writing, and completed matching to sample training designed to coordinate stimuli into classes of values, fused values, and neutral stimuli. Then, participants provided stimulus ratings of meaning and difficulty, and completed an approach and escape task. As DRR is defined in terms of mutual entailment, combinatorial entailment, and transformation of function (Hayes et al., 2001), these properties comprised the dependent variables.

Mutual and Combinatorial Entailment

All participants exhibited mutual entailment, and all but one exhibited combinatorial entailment, suggesting that individuals learn to coordinate events with values-relevant words quite readily. This is consistent with a robust literature on DRR (see Dymond et al., 2010; McLoughlin et al., 2020). The current study also sought to extend the literature, demonstrating entailment among a participant-specific stimulus and two arbitrary stimuli. This is only the second study (see Sandoz et al., 2020) to demonstrate DRR using participant-specific stimuli and the first in the context of values and values-affirmation. This is significant in terms of translation as values interventions involve expanding pre-existing relational networks including stimuli with which the participants have a long learning history (e.g., Wilson & Murrell, 2004), rather than building relations amongst novel, arbitrary stimuli.

Transformation of Function

The current study aimed to extend the literature on EC through DRR, with respect to both evaluative (e.g., Barnes-Holmes et al., 2000; Smyth et al., 2006) and consequential functions (e.g., Hayes et al., 1991; Valdivia-Salas et al., 2013; Whelan & Barnes-Holmes, 2004; Whelan et al., 2006). With respect to evaluative functions, transformation of function was assessed through (1) changes in ratings of meaningfulness and difficulty of arbitrary stimuli on a visual analog scale and (2) convergence of ratings of meaningfulness and difficulty between arbitrary and participant-generated stimuli. Both sets of analyses provided support for transformation of function. Participants rated the arbitrary stimulus they related to values words as significantly more meaningful than prior to relational training and rated the arbitrary stimulus they related to fused values words as significantly more difficult than prior to relational training. Both visual inspection and analyses of equivalence offered further support for transformation of evaluative functions With one exception, post-training ratings of meaningfulness and difficulty of combinatorially entailed stimuli were statistically equivalent to post-training ratings of participant-generated words. The meaningfulness rating of the combinatorially entailed valued stimulus was slightly attenuated relative to the participant generated stimulus, but the overall effect was still suggestive of functional equivalence.

With respect to consequential functions, transformation of function was assessed in terms of (1) divergence of approach/escape behavior between arbitrary stimuli related to different participant-generated stimuli and (2) convergence of approach/escape behavior between arbitrary and participant-generated stimuli. These analyses also provided support for transformation of function. Following relational training, participants approached the arbitrary stimulus they related to values words and escaped the arbitrary stimuli they related to fused value and neutral words. Statistical equivalence amongst entailed and participant generated stimuli was demonstrated for six of the nine tested pairings, with the remaining three showing slightly attenuated but still functionally equivalent mutually or combinatorially entailed escape/approach responses relative to the participant generated stimuli.

Despite evidence of transformation of function such that arbitrary stimuli were functionally equivalent to related participant-generated words, transformation was not uniform within nor across functions assessed. Meaningfulness increased for the derived values stimulus, but remained unchanged for the derived fused and neutral stimuli. Difficulty was unchanged for the derived values and neutral stimuli, but increased for the fused values stimulus. This suggests that, in addition to intentionally programmed contingencies, unprogrammed contextual cues for differential transformation of function (C_{func}) were also present. Part of the controlling context is likely the way that stimuli were generated and selected for inclusion in the conditional discrimination task. Participants were specifically directed to generate stimuli in three categories: (1) stimuli that represented their values, (2) stimuli that were not meaningful but were associated with guilt and shame, or (3) stimuli that were not meaningful. While stimuli were selected for inclusion empirically, most of the stimuli the algorithm selected were consistent with the experimental categories used for generation. In this way, the generation instructions may have made particular functions salient, just as an experimenter might provide oral instructions for participants to engage in one of two behaviors in response to each stimulus presentation (e.g., "I want you to look at that image and then I want you to either clap or wave your hands;" Roche et al., 2000). Then, the subsequent selection may have served to differentially reinforce the attribution of particular function to each stimulus (e.g., "Good, that is correct," or "No, that is wrong;" Roche et al., 2000). Future DRR research using pre-experimental functions might directly manipulate stimulus generation and selection procedures to examine their impact on transformation of function.

Limitations and Future Directions

As with any study, the conclusions that can be drawn from this study are limited by the particulars of its design – namely, stimulus generation, MTS structure, stimulus functions assessed, relations modeled, and focus on appetitive control. Consistent with previous research on stimulus generation (Sandoz & Hebert, 2015) participants generated value, fused value, and neutral words after a standard values-affirmation writing exercise (McQueen & Klein, 2006) with minimal guidance from the experimenter. In addition, a small subset of participants (*n* = 2; participants 11 and 13) rated unanticipated stimuli as highly meaningful, and therefore were assigned a value stimulus consisting of an experimenter-generated word (i.e., "address" and "machine"). As a result, some of the value stimuli (E1) generated appeared incongruent with common conceptualizations of values (e.g., "life" or "address;" Allport et al., 1960; Wilson et al., 2011). Visual analysis of these participants' responding revealed that both displayed substantial increases in meaningfulness ratings of the combinatorially entailed values stimulus (F1) from pre to post training (Figure 3). Further, both participants displayed a clear pattern of approach responses to the values stimulus (E1) and mutually entailed value stimulus (D1), with participant 11 fully approaching and participant 13 partially approaching the combinatorially entailed value stimulus (F1). Thus, these stimuli seemed to function consistent with values stimuli (i.e., rated as highly meaningful, evoked low rates of escape, and evoked high rates of approach) despite their intended "neutral" functions designated by the experimenters. This seemingly contradictory finding highlights the importance of directly assessing function of values stimuli at the participant level and cautions against presuming that the learning histories of participants will match researchers with regard to stimulus functions of potential value words. Future studies might build on this finding by explicitly examining whether degree of specificity of instructions or level of experiential intensity influences functional properties of the stimuli. Such studies could explore whether values that are not topographically congruent with common values conceptualizations (e.g., "life,") are functionally distinct from values stimuli that are congruent.

One noteworthy limitation of the present study was that all participants engaged in the values writing exercise as part of the idiographic stimulus generation in keeping with previous research (Sandoz & Hebert, 2015). Thus, it is difficult to determine with this design how much the values writing task contributed to the subsequent observations of mutual entailment, combinatorial entailment, and transformation of function. Future iterations of this study could differentiate the variability in patterns of DRR associated with the values writing by controlling for features of the writing (e.g., word count or key words), separating stimulus generation in time from the matching to sample portions of the study, or including other aspects of values stimulus generation in the literature that don't involve writing (e.g., picture selection; Sandoz & Hebert, 2015).

In addition, the MTS procedure can vary in structure of initial training among linear protocols (consequating relating each stimulus as both sample to one comparison and comparison to another sample), one-to-many (consequating relating a single sample to different comparisons, as in the current study), or many-to-one (consequating relating many samples to single comparison). Data have been mixed with regard to the relative effectiveness of these structures at establishing mutual and combinatorial entailment (e.g., Arntzen & Holth, 1997, 2000; Arntzen & Nikolaisen, 2011; Arntzen & Vaidya, 2008; Eilifsen & Arntzen, 2015; Fields et al., 1999; Hove, 2003; Saunders et al., 2005; Smeets & Barnes-Holmes, 2005). The present study data provide some support for extant literature on the appropriateness of one-to-many matching-to-sample procedures for establishing mutual and combinatorial entailment (e.g., Bordieri et al., 2016; Keenan et al., 2015; Stewart et al., 2015). Replications of this study might consider the use of many-to-one conditional discrimination procedures, which some studies have suggested is more effective (see Arntzen, 2012 for a review), and has been proposed to be more consistent with RFT (Barnes, 1994).

Stimulus functions were examined in terms of (1) ratings of meaningfulness and difficulty, and (2) responses consequated by increases or decreases of the size of the stimulus on the screen. In both cases, these behaviors are presumed to be part of a functional response class with socially significant behaviors such as those targeted in values-affirmation interventions (e.g., academic performance; Miyake et al., 2010). In the future, however, it might be useful to replicate this study using more direct measures of elicited functions that have been empirically associated with values contact (e.g., decreased cortisol response; Cresswell et al., 2005). In addition, it may be useful to consider both mutually and combinatorially entailed stimuli when assessing elicited functions as only functions of combinatorially entailed stimuli were assessed in this investigation. Finally, only evaluative functions were assessed prior to the MTS procedure. Future iterations of this study should include pre-training assessment of approach and escape behaviors.

Similarly, consistent with other computerized tasks assessing approach and escape, changing the size of the stimulus was cast in the instructions as “pushing away” or “pulling toward”, and interpreted as functionally equivalent to approach and escape. Although the orderliness in those data are consistent with those interpretations, it would be interesting to further examine stimulus functions using behaviors outside of the computer task (e.g., a stimulus preference assessment). This could be further extended to even more ecologically valid operant behaviors that have been empirically associated with values contact. For example, future examinations of transformation of function might include improved academic performance (e.g., enhanced scores on tests of numeracy and literacy; Cooke et al., 2012; Sherman, 2013), resilience to social ostracism (e.g., how quickly an ostracized person recovers their fundamental needs of self-esteem, meaningful existence, belonging, and control following their social exclusion experience; Burson et al., 2012; Williams, 2009).

In addition, this study modeled transformation of values functions using coordination relations, which are but a part of the complex hierarchical networks theoretically involved in values (Plumb et al., 2009; Villatte, 2020). Transformation of function across hierarchical relations has been demonstrated experimentally (Gil et al., 2012). One next step in investigating DRR involved in values could involve demonstrating transformation of values functions down a network from a superordinate value to a functional class of goals hierarchically related to that value, to several classes of specific behavioral steps hierarchically related to each valued goal. As another example, an experimental paradigm informed by relational density theory (Belisle & Dixon, 2020) could manipulate the density of values classes and explore the impact on approach and escape responding. Such an approach might allow for a direct assessment of broader patterns of approach behaviors and strengthen the link between basic RFT accounts and mid-level conceptualizations of values in ACT.

Further, the term ‘meaningful,’ while widely used in the literature, is rarely defined at a functional level in behavioral terms. It is generally assumed that the values writing task evokes meaningful stimuli for participants because it implicitly asks them to identify stimuli that are salient and appetitive to them. Future research in this area could build on a line of work on effects of meaningful stimuli on relational responding in the stimulus equivalence literature (e.g., Arntzen et al., 2018; de Almeida & de Rose, 2015; Tyndall et al., 2004). As Arntzen et al. (2018; p. 123-124) put it, “meaningless stimuli are those that do not have any specific discriminative functions, while meaningful stimuli bear some relation to other classes of stimuli.”

Finally, explicating values from a behavior analytic perspective has repeatedly focused on appetitive functions – how values establish reinforcers, increasing the likelihood of values-relevant behavior (e.g., Plumb et al., 2009). However, several studies report increased sensitivity to threat after contact with values, both in terms of attention bias (e.g., Klein & Harris, 2009) and emotional responsivity at the neurological level (e.g., Legault et al., 2012). A more complete analysis of how values interventions impact behavior might include comparing aversive and appetitive stimulation.

Conclusion

Despite limitations, this is the first study to experimentally model how it is that arbitrary events can come to reinforce important, life-changing behaviors through their relations with verbally constructed values. It is our hope that by offering an experimental account of this phenomenon, we create a foundation for continued scientific exploration of what values are, how they emerge, and how they are best intervened upon.

References

Allport, G. W., Vernon, P. E., & Linzey, G. (1960). *Study of values.* Houghton Mifflin.

Amd, M., & Roche, B. (2015). A derived transformation of valence functions across two 8-member comparative relational networks. *The Psychological Record, 65(3),* 523-540. https://doi.org/10.1007/s40732-015-0128-1

Amd, M., & Roche, B. (2016). A derived transformation of emotional functions using self-reports, implicit association tests, and frontal alpha asymmetries. *Learning & Behavior, 44*, 175-190. https://doi.org/10.3758/s13420-015-0198-6

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis*, *13*(1), 123-135. https://doi.org/10.1080/15021149.2012.11434412

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, *47*(2), 309-320. https://doi.org/10.1007/bf03395227

Arntzen, E., & Holth, P. (2000). Differential probabilities of equivalence outcome in individual subjects as a function of training structure. *The Psychological Record*, *50*(4), 603-628 . https://doi.org/10.1007/bf03395374

Arntzen, E., Nartey, R. K., & Fields, L. (2018). Graded delay, enhanced equivalence class formation, and meaning. *The Psychological Record, 68*(2), 123–140. https://doi.org/10.1007/s40732-018-0271-6

Arntzen, E., & Nikolaisen, S. (2011). Establishing equivalence classes in children using familiar and abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis*, *12*(1), 105-120.

https://doi.org/10.1080/15021149.2011.11434358

Arntzen, E., & Vaidya, M. (2008). The effect of baseline training structure on equivalence class formation in children. *Experimental Analysis of Human Behavior Bulletin*, *29*, 1-8.

Assaz, D. A., Roche, B., Kanter, J. W., & Oshiro, C. K. B. (2018). Cognitive defusion in Acceptance and Commitment Therapy: What are the basic processes of change? *The Psychological Record, 68,* 405-418*.* https://doi.org/10.1007/s40732-017-0254-z

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, *44*(1), 91-125.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. *The Psychological Record*, *50*(2), 251-265. https://doi.org/10.1007/bf03395355

Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. M. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record, 50*(3), 493-511. https://doi.org/10.1007/bf03395367

Belisle, J., & Dixon, M. R. (2020). Relational density theory: Nonlinearity of equivalence relating examined through higher-order volumetric-mass-density. *Perspectives on Behavior Science*, *43*(2), 259-283. https://doi.org/10.1007/s40614-020-00248-w

Bordieri, M. J., Kellum, K. K., Wilson, K. G., & Whiteman, K. C. (2016). Basic properties of coherence: Testing a core assumption of relational frame theory. *The Psychological Record, 66*(1), 83-98. https://doi.org/10.1007/s40732-015-0154-z

Burson, A., Crocker, J., & Mischkowski, D. (2012). Two types of value-affirmation: Implications for self-control following social exclusion. *Social Psychological and Personality Science,* *3*(4), 510-516. https://doi.org/10.1177/1948550611427773

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*(2), 215–224. https://doi.org/10.1177/0146167299025002007

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, *65*, 333-371. https://doi.org/10.1146/annurev-psych-010213-115137

Cooke, J. E., Purdie-Vaughans, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values-affirmation on identity development. *Journal of Personality and Social Psychology, 102*(3), 479-496. https://doi.org/10.1037/a0026312

Cresswell, J. D., Welch, W. T., Taylor, S. E., Sherman, D. K., Gruenewald, T. L., & Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress response. *Psychological Science, 16*(11)*,* 846-851. https://doi.org/10.1111/j.1467-9280.2005.01624.x

Dack, D., Reed, P., & McHugh, L. (2010). Multiple determinants of transfer of evaluative function after conditioning with free-operant schedules of reinforcement. *Learning & Behavior, 38*(4), 348-366. https://doi.org/10.3758/lb.38.4.348

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, 65*(3), 451–461. https://doi.org/10.1007/s40732-015-0120-9

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish journal of psychology*, *10*(02), 230-241. https://doi.org/10.1017/S1138741600006491

Dymond, S., Bennett, M., Boyle, S., Roche, B., & Schlund, M. (2019). Related to anxiety: Arbitrarily applicable relational responding and experimental research on fear and avoidance. *Perspectives on Behavior Science, 41*(1), 189-213. https://doi.org/10.1007/s40614-017-0133-6

Dymond, S., May, R. J., Munnelly, A., & Hoon, A. E. (2010). Evaluating the evidence base for relational frame theory: A citation analysis. *The Behavior Analyst, 33*(1), 97-117. https://doi.org/10.1007/bf03392206

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst, 23*(2), 239-254. https://doi.org/10.1007/bf03392013

Dymond, S., & Roche, B. (2013). *Advances in Relational Frame Theory: Research and Application*. Context Press.

Eilifsen, C., & Arntzen, E. (2015). Effects of training structure and the passage of time on trained and derived performance. *The Psychological Record, 65*(1), 1-12. https://doi.org/10.1007/s40732-014-0067-2

Fields, L., Hobbie, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record*, *49*(4), 703-723. https://doi.org/10.1007/bf03395336

Gil, E., Luciano, C., Ruiz, F. J., & Valdivia-Salas, S. (2012). A preliminary demonstration of transformation of functions through hierarchical relations. *International Journal of Psychology and Psychological Therapy*, *12*(1), 1-19.

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., Kerr, S., Tansey, L., Noel, P., Ferenbach, C., Masley, S., Roach, L., Lloyd, J., May, L., Clarke, S., & Remington, B. (2014). The development and initial validation of the Cognitive Fusion Questionnaire. *Behavior Therapy*, *45*(1), 83-101. https://doi.org/10.1016/j.beth.2013.09.001

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy*, *35*(4), 639-665. https://doi.org/10.1016/S0005-7894(04)80013-3

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* Kluwer Academic/Plenum Publishers.

Hayes, S. C., Kohlenberg, B., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental analysis of Behavior*, *56*(1), 119-137. https://doi.org/10.1901/jeab.1991.56-119

Hayes, S. C., Strosahl, K., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change.* Guilford Press.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390-421. https://doi.org/10.1037/a0018916

Hove, A. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*(4), 617-634. https://doi.org/10.1007/bf03395456

Keenan, M., Porter, I., & Gallagher, S. (2015). Merging separately established functional equivalence classes. *The Psychological Record, 65*(3), 435-450. https://doi.org/10.1007/s40732-015-0118-3

Kissi, A., Hughes, S., Mertens, G., Barnes-Holmes, D., De Houwer, J., & Crombz, G. (2017). A systematic review of pliance, tracking, and augmenting. *Behavior Modification, 41(5)*, 683-707. https://doi.org/10.1177/0145445517693811

Klein, W. M. P., & Harris, P. R. (2009). Self-affirmation enhances attentional bias toward threatening components of a persuasive message. *Psychological Science, 20*(12), 1463-1467. https://doi.org/10.1111/j.1467-9280.2009.02467.x

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355-362. https://doi.org/10.31234/osf.io/97gpc

Legault, L., Al-Khindi, T., & Inzlicht, M. (2012). Preserving integrity in the face of performance threat self-affirmation enhances neurophysiological responsiveness to errors. *Psychological Science, 23*(12), 1455-1460. https://doi.org/10.1177/0956797612448483

McCracken, L. M., Barker, E., & Chilcot, J. (2014). Decentering, rumination, cognitive defusion, and psychological flexibility in people with chronic pain. *Journal of Behavioral Medicine*, *37*(6), 1215-1225. https://doi.org/10.1007/s10865-014-9570-9

McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Convergence of multiple fields on a relational reasoning approach to cognition. *Intelligence*, *83*(Nov-Dec), 101491. https://doi.org/10.1016/j.intell.2020.101491

McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*(4), 289-354. https://doi.org/10.1080/15298860600805325

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values-affirmation. *Science*, *330*(6008), 1234-1237. https://doi.org/10.1126/science.1195996

Perez, W. F., Kovac, R., Nico, Y. C., Caro, D. M., Fidalgo, A. P., Linares, I., de Almeida, J. H., & de Rose, J. C. (2017). The transfer of Crel contextual control (same, opposite, less than, more than) through equivalence relations. *Journal of the Experimental Analysis of Behavior*, *108*(3), 318–334. https://doi.org/10.1002/jeab.284

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., Tusler, M., Wagner, L., & Crocker, J. (2017). Improving numeracy through values-affirmation enhances decision and STEM outcomes. *PLoS ONE 12*(7), e0180674. https://doi.org/10.1371/journal.pone.0180674

Plumb, J. C., Stewart, I., Dahl, J., & Lundgren, T. (2009). In search of meaning: Values in modern clinical behavior analysis. *The Behavior Analyst, 32*(1), 85-103. https://doi.org/10.1007/bf03392177

Prooijen, A. M., & Sparks, P. (2014). Attenuating initial beliefs: Increasing the acceptance of anthropogenic climate change information by reflecting on values. *Risk Analysis*, *34*(5), 929-936. https://doi.org/10.1111/risa.12152

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(2), 105–120. https://doi.org/10.1016/j.jbtep.2006.10.001

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes-Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record*, *50*(2), 267-291. https://doi.org/10.1901/jeab.1997.67-275

Sandoz, E. K., & Anderson, R. (2015). Building awareness, openness, and action: Values work in psychotherapy. *The Behavior Therapist*, *38*(3), 60-70.

Sandoz, E. K., Bordieri, M. J., Tyndall, I., & Auzenne, J. (2020). A preliminary examination of derived relational responding in the context of body image. *The Psychological Record*, *71*(2)*,* 1-16. https://doi.org/10.1007/s40732-020-00439-6

Sandoz, E. K., & Hebert, E. R. (2015). Meaningful, reminiscent, and evocative: An initial examination of four methods of selecting idiographic values-relevant stimuli. *Journal of Contextual Behavioral Science*, *4*(4), 277-280. https://doi.org/10.1016/j.jcbs.2015.09.001

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record*, *55*(4), 539-559. https://doi.org/10.1007/bf03395526

Sherman, D. K. (2013). Self-affirmation: Understanding the effects. *Social and Personality Psychology Compass, 7*(11), 834-845. https://doi.org/10.1111/spc3.12072

Sherman, D. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality & Social Psychological Bulletin, 26*(9), 1046-1058. https://doi.org/10.1177/01461672002611003

Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes*, *69*(3), 281-293. https://doi.org/10.1016/j.beproc.2004.12.009

Smyth, S., Barnes-Holmes, D., & Forsyth, J. P. (2006). A derived transfer of simple discrimination and self-reported arousal functions in spider fearful and non-spider-fearful participants. *Journal of the Experimental Analysis of Behavior, 85*(2), 223-246. https://doi.org/10.1901/jeab.2006.02-05

Stewart, I., Hooper, N., Walsh, P., O'Keefe, R., Joyce, R., & McHugh, L. (2015). Transformation of though suppression functions via same and opposite relations. *The Psychological Record*, *65*, 375-399. https://doi.org/10.1007/s40732-014-0113-0

Tesser, A., Martin, L. L., & Cornell, D. P. (1996). On the substitutability of self-protective mechanisms. In P. M. Golwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 48-68). Guilford.

Tyndall, I. T., Roche, B., & James, J. E. (2004). The relation between stimulus function and equivalence class formation. *Journal of the Experimental Analysis of Behavior, 81*(3), 257-266. https://doi.org/10.1901/jeab.2004.81-257

Valdivia-Salas, S., Dougher, M. J., & Luciano, C. (2013). Derived relations and generalized alteration of preferences. *Learning and Behavior*, *41*, 205-217. https://doi.org/10.3758/s13420-012-0098-y

Villatte, M. (2020). Using clinical RFT to enhance our ACT interventions: The example of values work. In M. E Levin, M. P. Twohig, and J. Krafft (Eds.) *Innovations in Acceptance and Commitment Therapy: Clinical Advancements and Applications in ACT (*pp. 30-40). Context Press/New Harbinger.

Whelan, R., & Barnes-Holmes, D. (2004). The transformation of consequential functions in accordance with the relational frames of same and opposite. *Journal of the Experimental Analysis of Behavior, 82,* 177-195. https://doi.org/10.1901/jeab.2004.82-177

Whelan, R., Barnes-Holmes, D., & Dymond, S. (2006). The transformation of consequential functions in accordance with the relational frames of more-than and less-than. *Journal of the Experimental Analysis of Behavior*, *86*(3), 317-335. https://doi.org/10.1901/jeab.2006.113-04

Williams, K. D. (2009). Ostracism: A temporal need-threat model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol 41, pp. 275-314). Elsevier Academic Press.

Wilson, K. G., & Murrell, A. R. (2004). Values work in acceptance and commitment therapy: Setting a course for behavioral treatment. In S. C. Hayes, V. M. Follette, & M. Linehan (Eds.), *Mindfulness & Acceptance: Expanding the Cognitive-Behavioral Tradition* (pp. 120-151). Guilford Press.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2011). The Valued Living Questionnaire: Defining and measuring valued action within a behavioral framework. *The Psychological Record*, *60*(2), 249-272. https://doi.org/10.1007/bf03395706

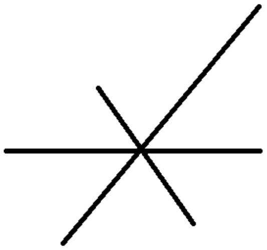
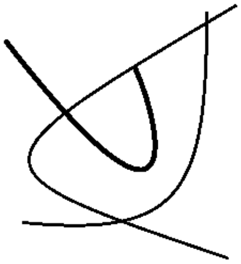
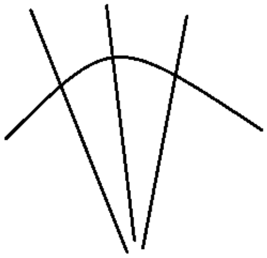
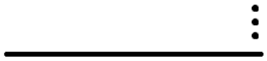
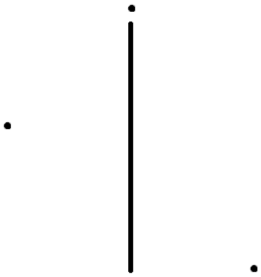
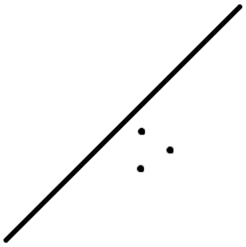
	1	2	3
D			
E	**Value**	**Fused** **Value**	**Neutral**
F			

Figure 1. Stimuli Used in Relational Training and Testing

Note. E and F stimuli varied for each participant based on their ratings in Phase 3. A total of six F stimuli were rated by participants with three representative samples displayed above.

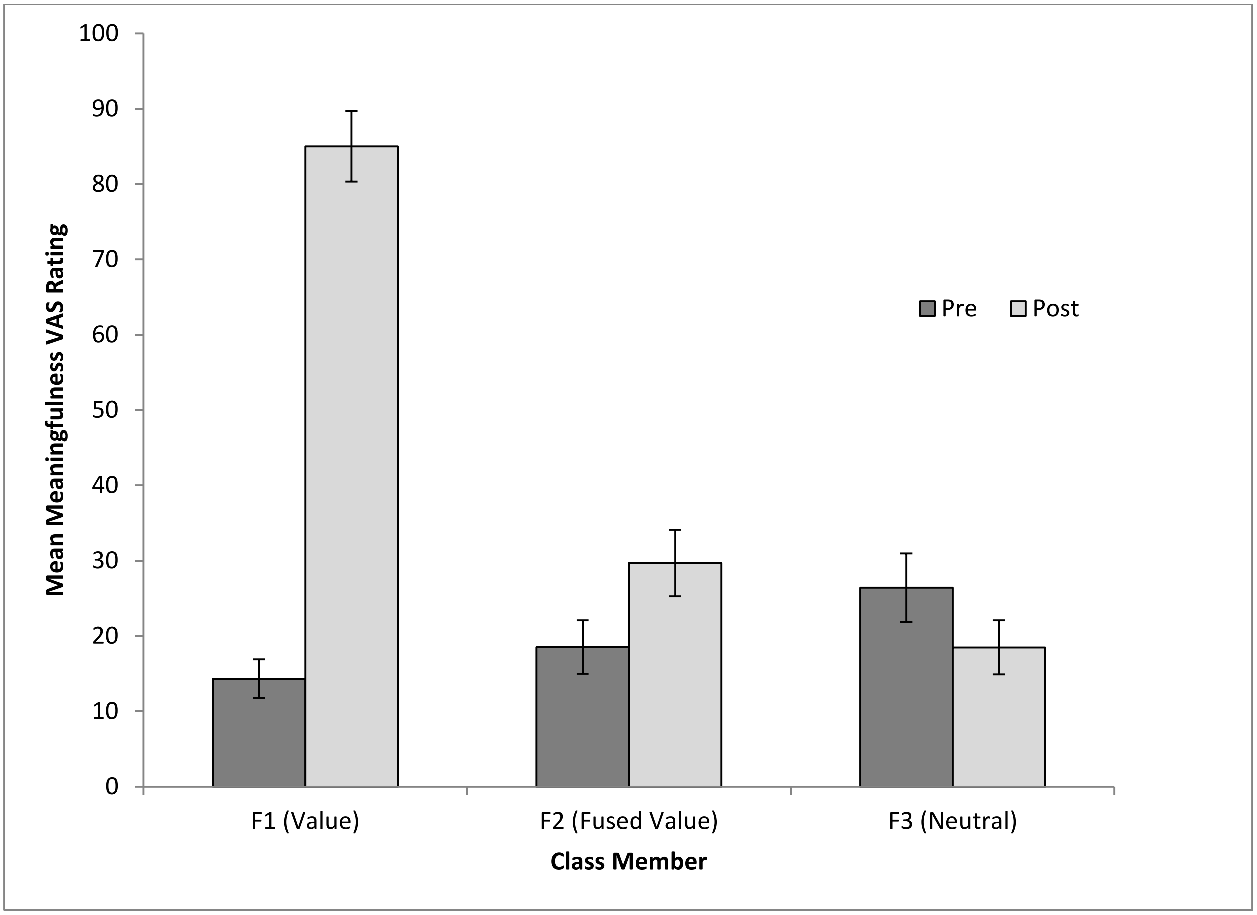
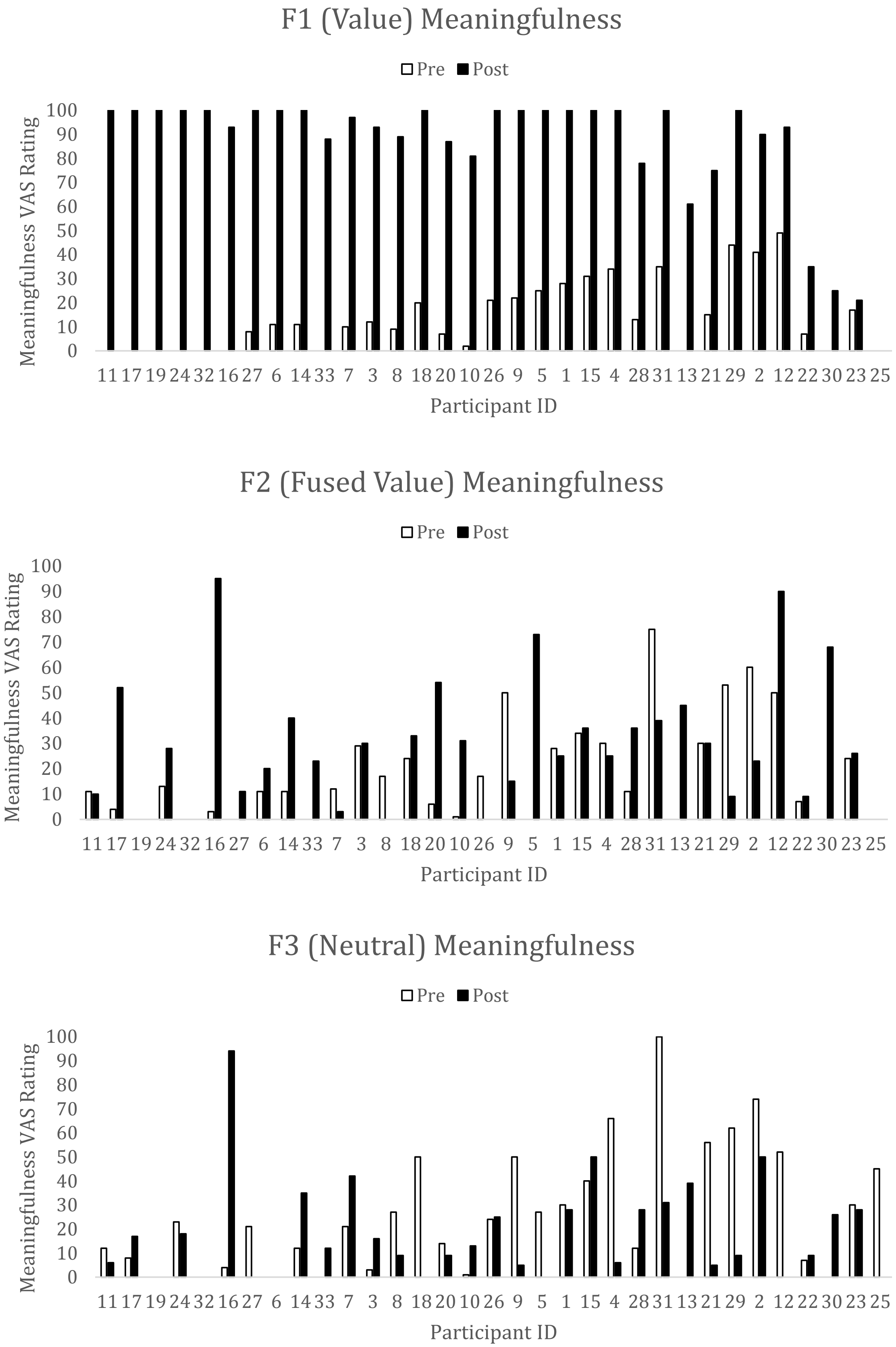


Figure 2. Group level differences in meaningfulness VAS ratings from pre to post class acquisition training across combinatorially entailed stimuli.


F1 (Value) Meaningfulness

F2 (Fused Value) Meaningfulness

F3 (Neutral) Meaningfulness

Figure 3. Participant level differences in meaningfulness VAS ratings from pre to post class acquisition training across combinatorially entailed stimuli

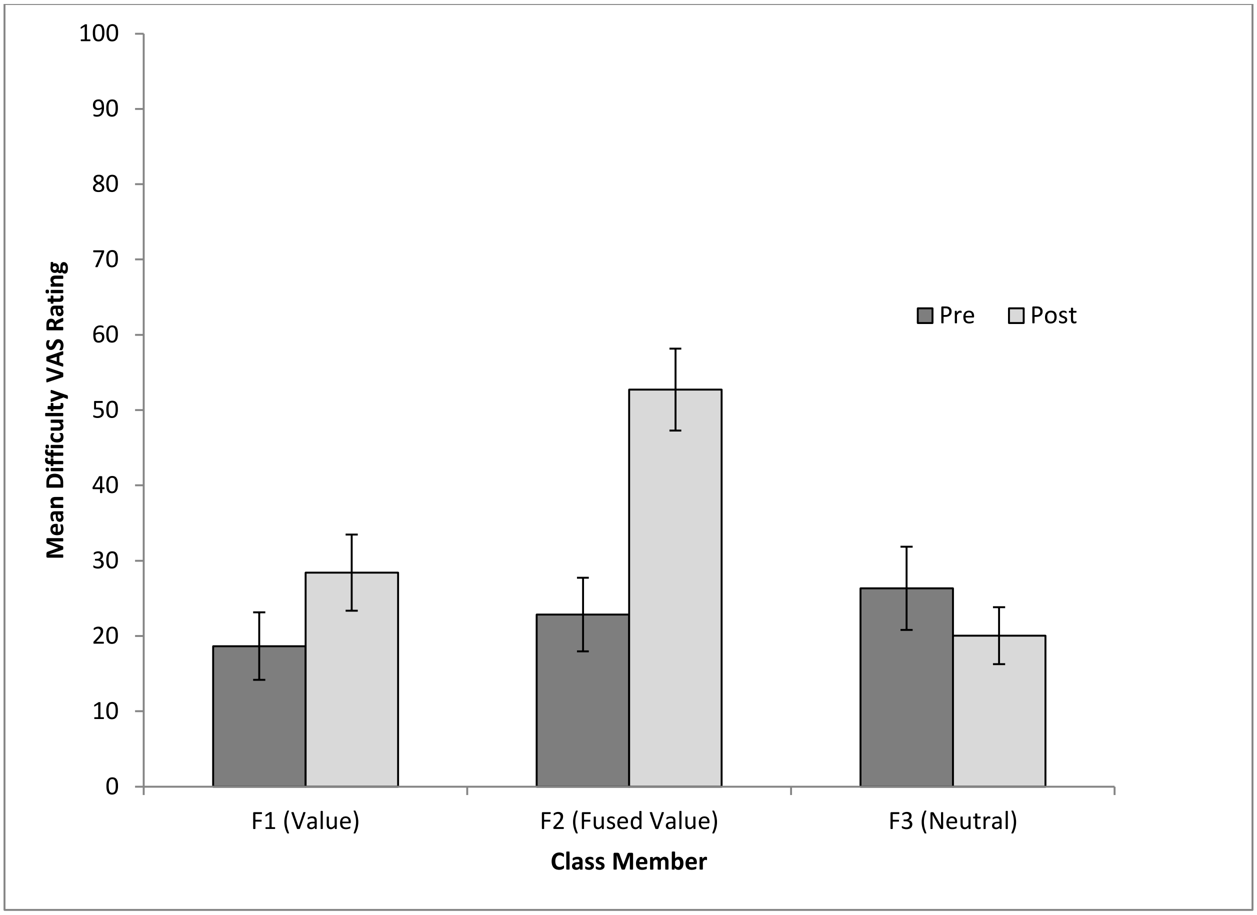
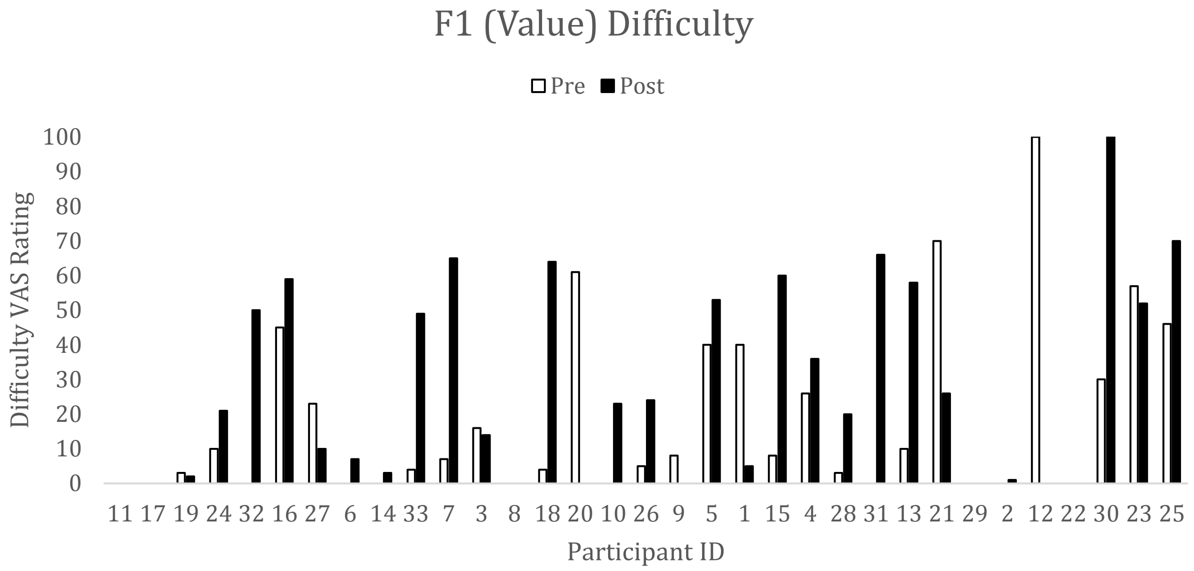
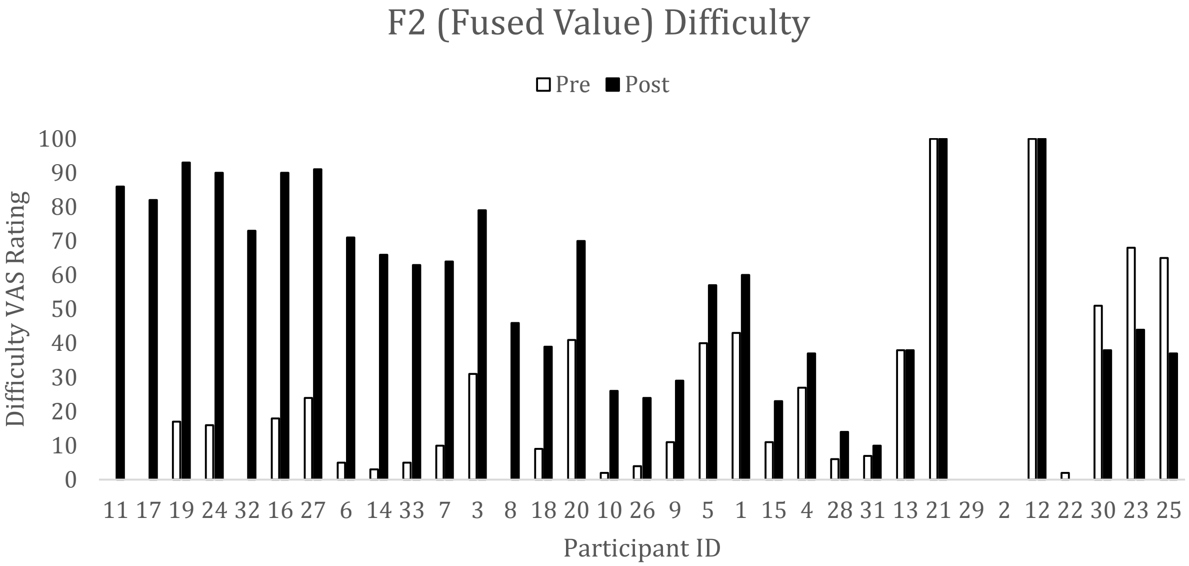


Figure 4. Group level differences in difficulty VAS ratings from pre to post class acquisition training across combinatorially entailed stimuli.


F1 (Value) Difficulty


F2 (Fused Value) Difficulty

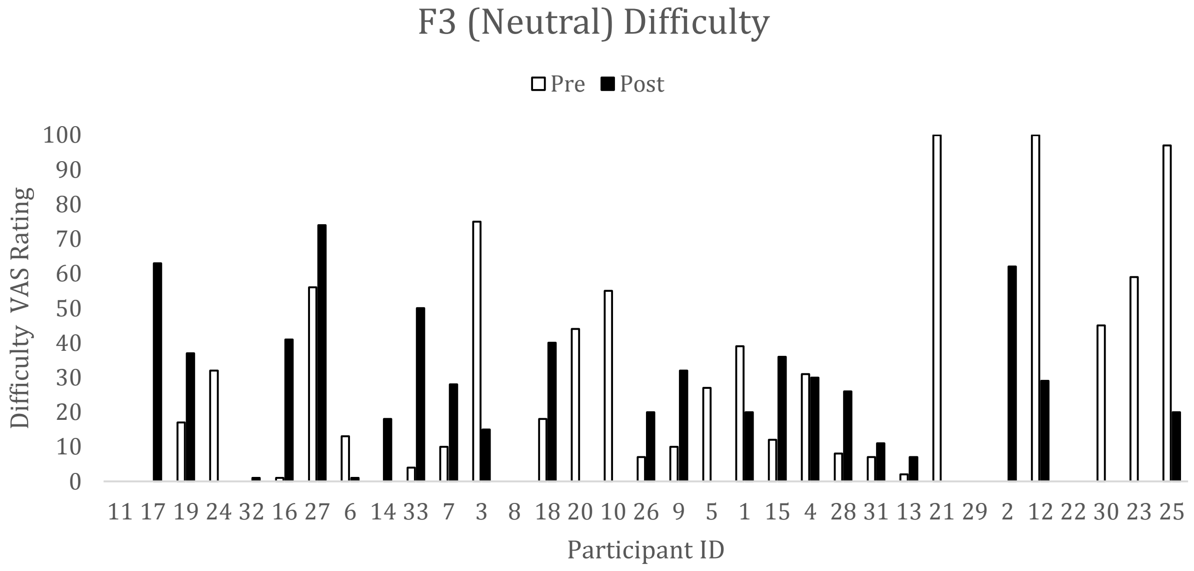

F3 (Neutral) Difficulty

Figure 5. Participant level differences in difficulty VAS ratings from pre to post class acquisition training across combinatorially entailed stimuli.

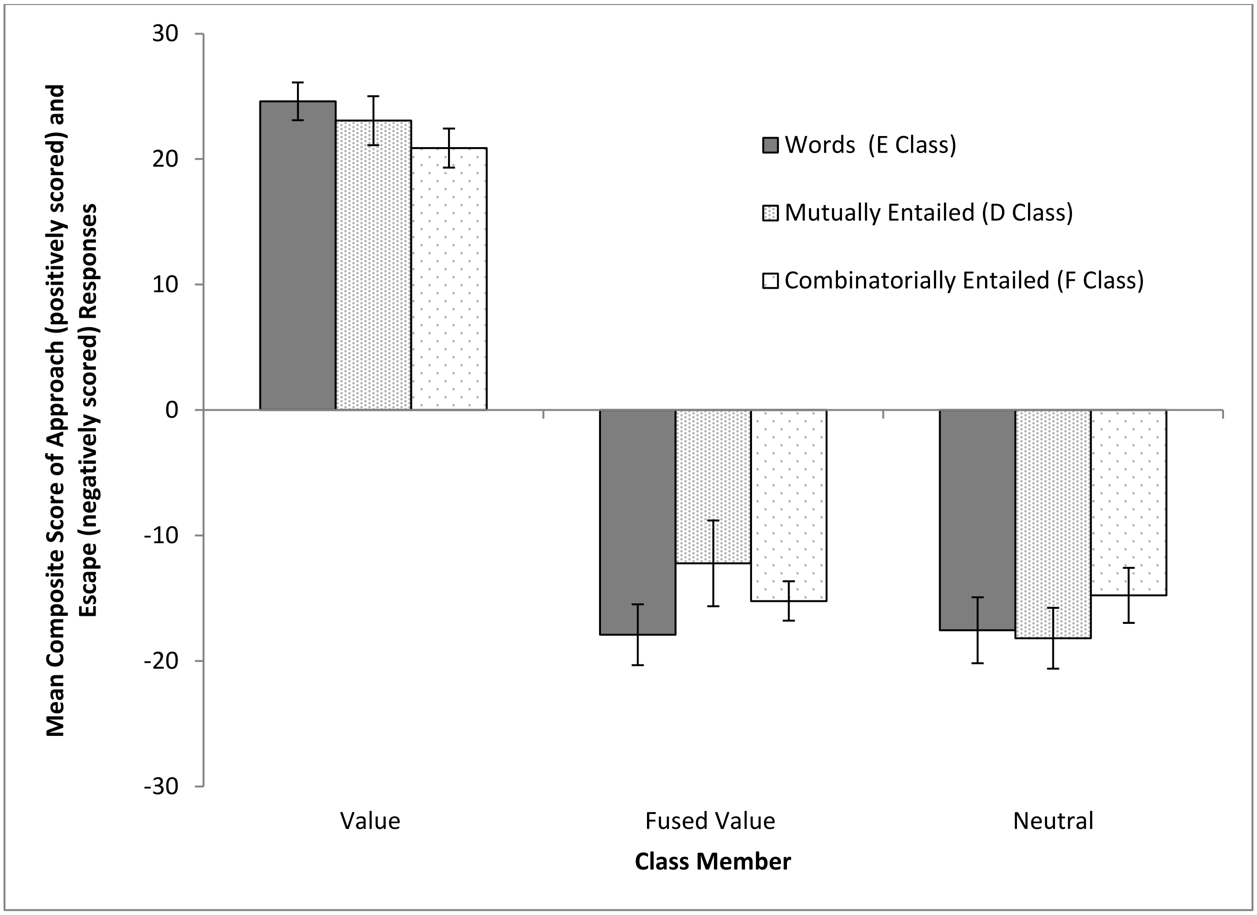
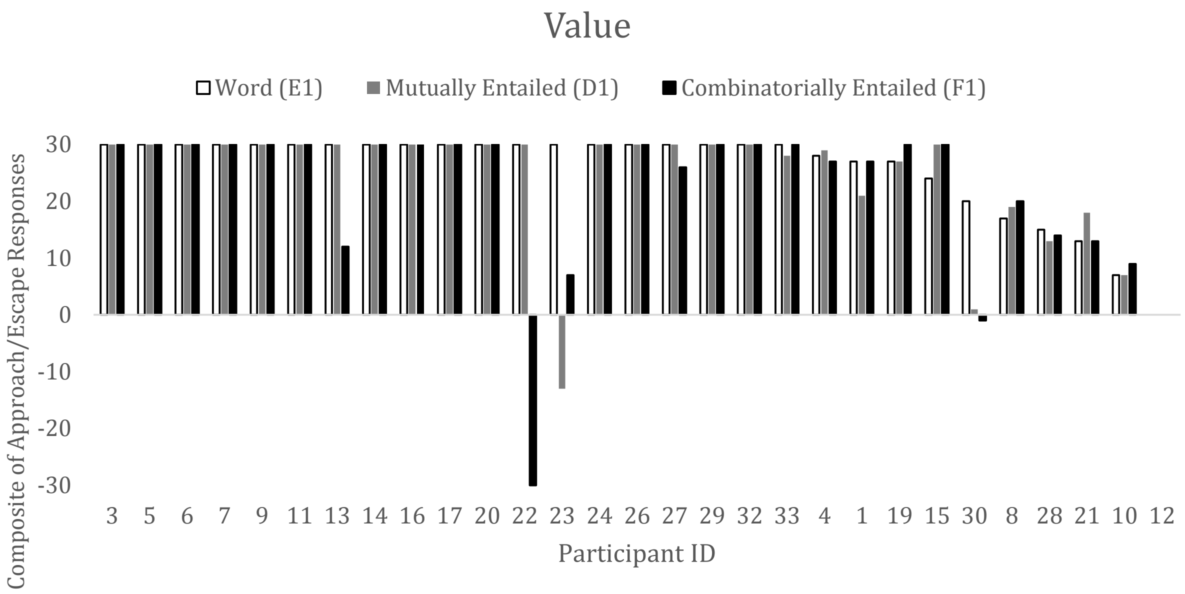
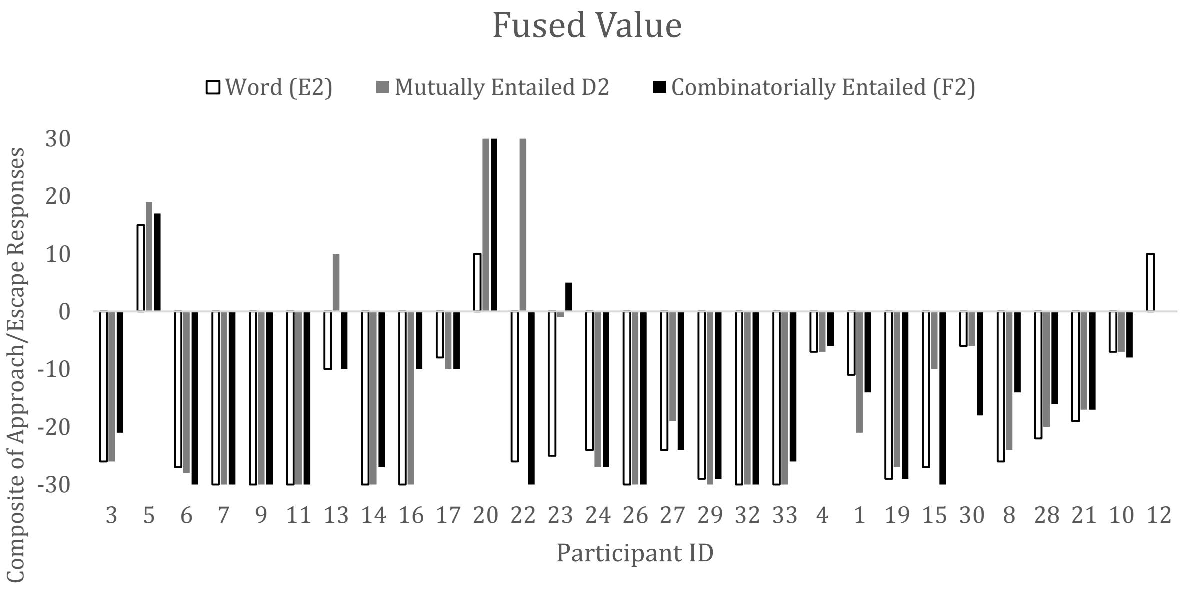


Figure 6. Group level composite score of approach (positively scored) and escape (negatively scored) responses to study stimuli.


Value
□ Word (E1)
■ Mutually Entailed (D1)
■ Combinatorially Entailed (F1)
Composite of Approach/Escape Responses
30
20
10
0
-10
-20
-30
3 5 6 7 9 11 13 14 16 17 20 22 23 24 26 27 29 32 33 4 1 19 15 30 8 28 21 10 12
Participant ID


Fused Value
□ Word (E2)
■ Mutually Entailed D2
■ Combinatorially Entailed (F2)
Composite of Approach/Escape Responses
30
20
10
0
-10
-20
-30
3 5 6 7 9 11 13 14 16 17 20 22 23 24 26 27 29 32 33 4 1 19 15 30 8 28 21 10 12
Participant ID

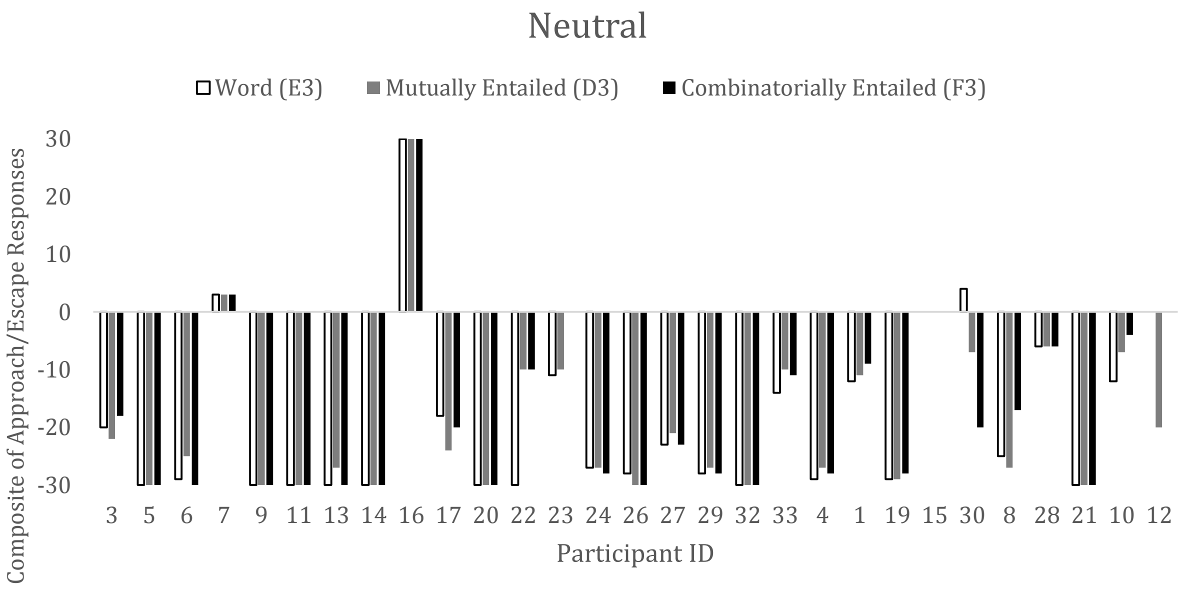

Neutral
□ Word (E3)
■ Mutually Entailed (D3)
■ Combinatorially Entailed (F3)
Composite of Approach/Escape Responses
30
20
10
0
-10
-20
-30
3 5 6 7 9 11 13 14 16 17 20 22 23 24 26 27 29 32 33 4 1 19 15 30 8 28 21 10 12
Participant ID

Figure 7. Participant level composite score of approach (positively scored) and escape (negatively scored) responses to study stimuli.

Table 1. *Descriptive Statistics of Class Acquisition Performance*

Variable	Mean	SD	Median	Min	Max
Trial Blocks to Criterion					
Train D-E	2.00	1.04	2	1	5
Train D-F	2.53	1.33	2	1	7
Mixed Train D-E/D-F	1.18	0.58	1	1	4
Total	5.71	2.47	5	3	15
Training Time (minutes)	10:49	5:18	9:06	6:27	34:20
Testing Accuracy (% correct)					
Mutual Entailment	98.09	3.12	100	89	100
Combinatorial Entailment	97.18	5.88	100	72	100

Note. Data from the second testing blocks were used for participants who received remedial mixed training.

Table 2. *Equivalence Analysis of Meaningfulness and Difficulty Ratings between Direct and Combinatorially Entailed Stimuli Post-Class Acquisition Training*

			Student’s t-test (NHST)		Equivalence t-test (TOST)	
Relation	E Stimulus Mean Rating	F Stimulus Mean Rating	t	p	t	p
Meaningfulness						
E1-F1 (Value)	91.15 (18.45)	85.03 (26.44)	1.61	.118	-1.26	.108
E2-F2 (Fused Value)	28.64 (25.02)	29.67 (25.02)	-0.36	.724	2.52	.009*
E3-F3 (Neutral)	17.67 (23.94)	18.48 (20.40)	-0.19	.854	2.69	.006*
Difficulty						
E1-F1 (Value)	28.73 (27.98)	28.42 (28.55)	0.09	.928	-2.78	.005*
E2-F2 (Fused Value)	52.76 (31.04)	52.73 (30.82)	0.01	.993	-2.86	.004*
E3-F3 (Neutral)	17.15 (21.88)	20.03 (21.39)	-1.09	.286	1.79	.042*

*p < .05

NHST = Null-hypothesis significance testing; TOST = Two one-sided tests equivalence testing

Table 3. *Equivalence Analysis of Approach/Escape Composite Scores between Study Stimuli*

Relation	1st Stimulus Mean Composite	2nd Stimulus Mean Composite	Student's t-test (NHST)		Equivalence t-test (TOST)	
			t	p	t	p
Mutually Entailed						
E1-D1 (Value)	25.79 (7.92)	23.79 (11.42)	1.21	.235	-1.75	.046*
E2-D2 (Fused Value)	-19.24 (13.49)	-14.86 (18.20)	-1.86	.074	1.11	.139
E3-D3 (Neutral)	-18.76 (14.83)	-18.76 (13.98)	0.00	1.00	-2.96	.003*
F1-D1 (Value)	21.86 (14.03)	23.79 (11.42)	-0.84	.410	2.13	.021*
F2-D2 (Fused Value)	-17.03 (15.23)	-14.86 (18.20)	-0.87	.393	2.10	.023*
F3-D3 (Neutral)	-18.17 (14.47)	-18.76 (13.98)	0.71	.481	-2.25	.016*
Combinatorially Entailed						
E1-F1 (Value)	25.79 (7.92)	21.86 (14.03)	1.67	.107	-1.30	.103
E2-F2 (Fused Value)	-19.24 (13.49)	-17.03 (15.23)	-1.37	.181	1.59	.062
E3-F3 (Neutral)	-18.76 (14.83)	-18.17 (14.47)	-0.46	.653	2.51	.009*

*p < .05

NHST = Null-hypothesis significance testing; TOST = Two one-sided tests equivalence testing

Conflict of Interest

Compliance with Ethical Standards

Funding

This research was funded, in part, by a Research Competitiveness Subprogram Grant LEQSF(2011-14)-RD-A-29 through the Louisiana Board of Regents.

Conflict of Interest

Given their role as Editorial Board Members, neither Mike Bordieri nor Ian Tyndall had no involvement in the peer-review of this article and had no access to information regarding its peer-review.

Given their role as Associate Editor, Emily Sandoz had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to Dr. Mike Levin.

Authors declare no other conflicts of interest.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.