A|B|A|I
Association for Behavior Analysis International

Check for updates

Tailoring of Painful Stimuli Used for Exploring Transfer of Function

Jon Magnus Eilertsen[1] · **Erik Arntzen**[1]

Abstract

Fifteen college students rated the degree of painfulness of six images showing needle injections to different parts of a human hand. The images were rated on a scale ranging from 1 (not painful) to 5 (severely painful). Next, the participants were trained six conditional discriminations with abstract shapes as stimuli and tested for the formation of three 3-member equivalence classes in a one-to-many (OTM) training structure (AB/AC). The test for emergent relations were followed by training a novel set of stimuli (D) to the A stimuli. The needle injection images rated with the highest and lowest perceived degree of painfulness were used as stimulus D1 and D2, respectively. An image where the needle was replaced with a Q tip was used as stimulus D3 and assumed to be perceived as not painful. The DA training was followed by a matching-to-sample (MTS) test for the formation of three 4-member classes, and the participants were asked to choose one out of three identical bottles with water labelled with the B stimuli. Finally, the participants were asked to rate the degree of painfulness of the B stimuli. The main findings were that the B and D stimuli were not rated significantly different, indicating a transfer of function in the ratings from D to B stimuli. In addition, participants avoided choosing bottle B1 over B2 and B3, but they did not differentiate between B2 and B3.

Keywords Stimulus equivalence · Transfer of function · Avoidance · Painful stimuli · Class expansion

Equivalence classes consist of an *n*-number of mutually interchangeable stimuli. For example, training six conditional discriminations (A1B1, A2B2, A3B3, B1C1, B2C2, and B3C3) might result in the emergence of three 3-member classes (ABC). If the classes inherit the properties of reflexivity (AA, BB, and CC), symmetry (BA and CB), and transitivity (AC), they are stimulus equivalence classes. A combined equivalence test (CA) simultaneously tests for symmetry and transitivity (Sidman & Tailby, 1982). Equivalence classes can be expanded by training members of an existing class to members of another existing class, thus merging the classes. For example, in a study by Sidman, Kirk, and Willson-Morris (1985) participants were first trained and tested for the formation of two separate three 3-member classes (ABC, DEF) in a one-to-many (OTM) training structure (AB, AC). Next, participants were trained three more conditional discriminations,

E1C1, E2C2, and E3C3, and finally tested for the emergence of three 6-member classes. The results showed that training 15 conditional relations led to the emergence of 60 relations.

Expanding equivalence classes as in Sidman et al. (1985) show the potential of mimicking how relations among stimuli emerge in real life. Also, if a member of an equivalence class is given a specific function, that function might transfer to the remaining members of a class (Fields & Garruto, 2009). The emergence of stimulus functions has been referred to as transfer of function (ToF; Dougher & Markham, 1996; Dymond & Rehfeldt, 2000). In one experiment studying ToF, Hayes, Devany, Kohlenberg, Brownstein, and Shelby (1987) trained six participants four conditional discriminations (A1B1, A1C1, A2B2, and A2C2) and tested for the emergence of two 3-member equivalence classes. For three of the participants, clapping was reinforced in the presence of stimulus B1 and waving was reinforced in the presence of B2. In a subsequent transfer test, they found that C1 occasioned clapping and C2 occasioned waving, showing that the discriminative control of a stimulus could transfer to other members of an equivalence class.

In another experiment, Dougher, Augustson, Markham, Greenway, and Wulfert (1994) provide an example of ToF with respondent eliciting functions measured with skin conductance. Participants trained and tested for the emergence of

✉ Jon Magnus Eilertsen
   jonmagnuseilertsen@gmail.com

   Erik Arntzen
   erik.arntzen@equivalence.net

[1] Department of Behavioral Science, Oslo Metropolitan University, Box 4 St. Olavs Plass, 0130 Oslo, PO, Norway

two 4-member equivalence classes (ABCD). Stimulus B1 was subsequently paired with mild electric shock and B2 was presented without shock. In the following transfer test, the eliciting functions were transferred to the C and D stimuli for six out of eight participants. A respondent extinction procedure for one of the class members resulted in the respondent extinction for the remaining members of the equivalence classes.

Research on derived stimulus relations and the ToF have also contributed to a behavioral explanation of clinically relevant behaviors such as fear, avoidance, and anxiety (e.g., Dougher, Twohig, & Madden, 2014; Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015; Dymond et al., 2011; Friman, Hayes, & Wilson, 1998; Lewon & Hayes, 2014). For example, in the study by Augustson and Dougher (1997) eight participants trained and tested for the emergence of two 4-member classes with an OTM training structure (A1B1C1D1, A2B2C2D2). Then, a classical conditioning procedure was introduced to establish B1 as CS+ and B2 as CS-. In Phase 3, the participants learned an avoidance response by pressing a telegraph key with an FR 20 schedule. The key-pressing response eliminated the presentation of B1 and subsequent shock presentation. If the participants failed to produce the avoidance response, shock was presented. In Phase 4, the participants were tested for the transfer of the avoidance response to C and D stimuli together with the presentation of the B stimuli. Finally, in Phase 5, participants were exposed to an equivalence test to assess if the classes were still intact. All eight participants successfully emitted avoidance responding to B1, C1, and D1.

A variety of experimental arrangements have been employed to evaluate the influence of preference for stimuli in equivalence classes (e.g., Arntzen, Eilertsen, & Fagerstrøm, 2016a; Arntzen, Fagerstrøm, & Foxall, 2016b; Barnes-Holmes, Keane, Barnes-Holmes, & Smeets, 2000; Bortoloti & de Rose, 2009). For example, Arntzen et al. (2016b) trained participants six conditional discriminations in an OTM training structure (AB/AC) resulting in the emergence of three 3-member equivalence classes. The classes were expanded by training three new stimuli (D1, D2, D3) to the A stimuli (A1, A2, A3). The D stimuli consisted of a happy (D1), a neutral (D2), and a sour smiley face (D3). The B stimuli (B1, B2, and B3) were attached onto three identical water bottles. When the participants had formed the three 4-member equivalence classes, they were asked to choose one out of the three bottles of water with label B1, B2, and B3, respectively. Thirteen out of the 16 participants chose the bottle labeled with the B1 stimulus, which was in the same equivalence class as the smiley face (D1), showing preference towards the bottle labelled with stimulus B1. The findings have been replicated with the use of different abstract stimuli, different types of D stimuli, and with a control group (Arntzen et al., 2016a; Eilertsen & Arntzen, 2017).

The transfer of avoidance behavior related to aversive stimuli has not been studied with the experimental arrangements utilized by Arntzen et al. (2016b). Therefore, in the current experiment, we wanted to explore if the preference test could also be employed as a choice test to evaluate avoidance of stimuli. This could be done by using D stimuli containing perceived aversive emotional functions and then test if participants avoid choosing the bottles labeled with the stimuli in the same equivalence class as the most aversive D stimuli.

To our knowledge, the ToF studies have not tailored the stimuli to each participant, but rather used stimuli defined by the experimenters. Hence, conflicting stimulus control and lack of coherence between the stimulus control intended by the experimenter and the actual stimulus control generated by the contingencies might account for some of the variability in some studies of the ToF (see McIlvane & Dube, 2003 for a description of stimulus control topography). For instance, participants might have a high degree of interindividual variance in how stimuli are perceived.

The purpose of the current study was to explore if a stimulus evaluation procedure could be used to select stimuli for each participant based on their ratings, in this way tailoring the stimuli for each participant. Next, we wanted to see if the specific ratings could transfer to other members of an equivalence class. Finally, we wanted to study if participants avoided choosing bottles labelled with stimuli in the same equivalence class as images with different perceived painfulness.

Method

Participants

Fifteen participants, four men and 11 females, with the mean age of 26.5 years, participated with the chance to win a 5,000 Norwegian kroner (approximately $ 542) universal gift card. Three more participants were dismissed from further experimentation. One due to a procedural error, and two did not respond in accordance with equivalence. Participants were recruited at Oslo Metropolitan University and by personal contacts. None of the participants had any prior knowledge about stimulus equivalence. All participants were handed a consent form upon arriving, which contained general information about the experimental setting. Furthermore, the form contained information about participant anonymity, and that they could withdraw from the experiment at any given time. The participants were asked to sign the form before proceeding with the experimental phases. All the participants were thanked and debriefed at the end of the experiment, and the results were explained to the participants. The focus of the debriefing was to answer any questions the participants had regarding the experiment, and to provide some insight into stimulus equivalence research.

Apparatus and Setting

The experiment was conducted in two 200 cm x 135 cm cubicles located in a larger quiet room. The cubicles were furnished with a table and a chair. An HP ProBook 470 GP laptop computer running on a Windows 10 64-bit operative system with a 17.4-inch screen and an external mouse was used to run a customized software program. The program administered the presentation of stimuli in the conditional discrimination training and testing.

Stimuli

Figure 1 shows the needle injection images presented to the participants in Phase 1 (see description of the phases, below). The images were selected out of a set of 32 images that were evaluated by 115 participants in a previous study—indicating the perceived painfulness of the needle injections (Lamm, Meltzoff, & Decety, 2010).[1] Figure 2 displays the stimuli used for the conditional-discrimination training and testing. The stimuli consisted of abstract shapes shown as the A, B, and C stimulus sets. The figure includes two out of the six needle injection images used as the D1 and D2 stimuli. Stimulus D3 displays a Q tip being pressed against the index finger knuckle of a left hand. The D3 stimulus remained the same for all the participants. The size of the stimuli varied from 1.8 cm to 3.7 cm in height and from 1.2 cm to 2.2 cm in width. The three bottles used during the choice test in Phase 6 were identical blank bottles of water with screw caps (see Figure 3). Each of the three bottles had a printout of the B stimuli attached to them (B1, B2, and B3). The stimulus printouts were 9.3 cm in height and 5 cm in width. The B stimuli displayed on the printouts varied from 3.5–4.3 cm in height and 1.4–2 cm in width. The bottles were lined up on a table approximately 4 cm between them.

Design, Independent and Dependent Variables

The current experiment employed a one-group pretest/posttest design with a tailoring of stimuli, expansion of equivalence classes, and a preference test. The independent variable is the tailored stimuli implemented in the conditional discrimination training and testing for the emergence of equivalence classes. The dependent variables are primarily the choices in the choice test and the evaluation of the B stimuli, trials to criterion, and equivalence class formation (symmetry, transitivity, and equivalence).

---

[1] Necessary permissions to use the images have been obtained by the authors

Procedure

The procedure consisted of seven phases: (1) rating of needle injections, (2) training of conditional discriminations, (3) test of emergent relations, (4) class expansion training with three new conditional discriminations, (5) test of emergent relations, (6) choice test, and (7) rating of the B stimuli.

Phase 1: Baseline evaluations of images of needle injections After the participants had signed the consent forms, they were handed six sheets containing images of needle injections towards different areas of a human hand. At the bottom of each image there was a 5-point Likert type scale. An instruction was presented in the top of each sheet, above the image. The instruction stated, "Evaluate the picture on the scale below. Insert an X in the box that fits best to how you experience the image." The five boxes were placed under the images and were numbered from 1 to 5 where statements under the boxes were as follows: (1) not painful, (2) slightly painful, (3) moderately painful, (5) very painful, and (5) severely painful. The participants were asked to "rate each of the three images on the corresponding scale." The order of the sheets with the images were shuffled for each participant to minimize the chance of any order effect. The participants were told that they could flip through and observe all the images before they rated them. When the participants rated the images, the experimenter left the room, so as not to influence the rating. When the participants had rated the images, the experimenter gathered the sheets. The image rated as the most painful out of the six images was selected to be used as stimulus D1 and the image rated as the least painful was selected to be used as stimulus D2 during the class expansion training in Phase 4. If two or more images were rated with similar degree of painfulness, for example, if three images were given the rating of 5 points "severely painful" one of the images was selected by using the randomization application Random Number Generator by UX Apps (2017) for Android and selected as stimulus D1. The same was done for the selection of stimulus D2. The individually tailored D stimuli were employed during Phases 4 and 5.

Phase 2: Conditional discrimination training The participants were seated in the cubicle in front of the laptop computer. The computer screen displayed the following instructions:
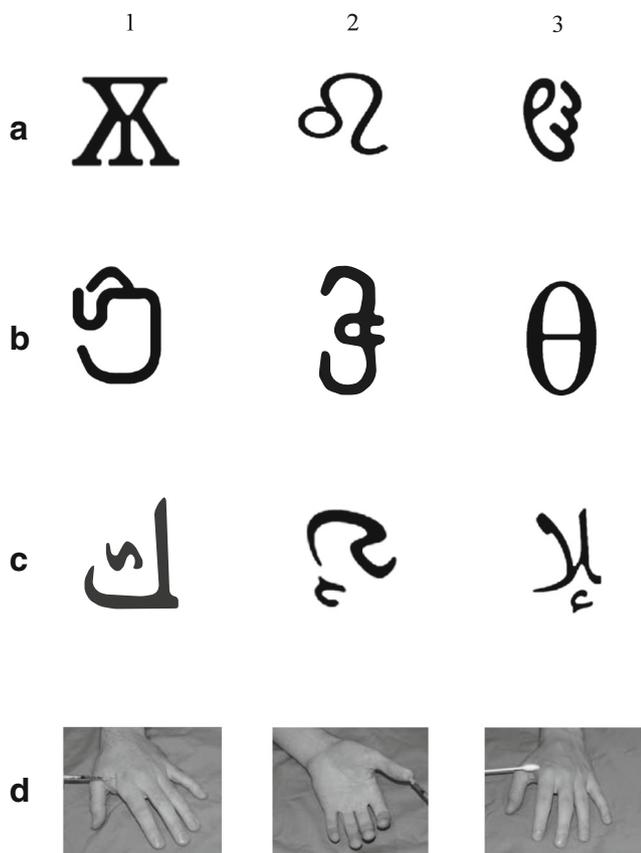
> Thank you for participating in this experiment. This is an experiment within learning psychology and requires no prior computer-knowledge. In short, you should click some stimuli that appear on the screen. The goal is to get as many correct as possible. When you move the mouse cursor on the stimulus in the middle and click it, more stimuli will appear on the screen. Mouse clicks on the

Fig. 1 Shows the six images presented to the participants in Phase 1 of the experiment. The image rated as most and least painful were employed as stimulus D1 and, D2 respectively in Phase 4.
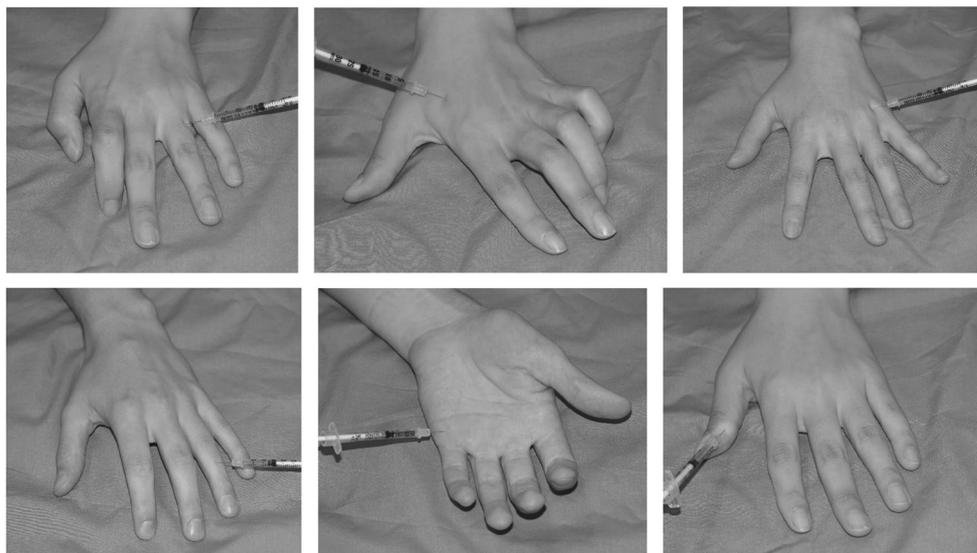
correct ones in the corners will be followed by the text "Correct" or similar on the screen. Clicking on one of the wrong ones, will be followed by the text "Wrong". That is how you find out what is right and wrong. After a

Fig. 2 Gives an overview of the stimuli used in the conditional discrimination training and testing. Note that the Dl and D2 stimuli were different for each participant based on how they rated the images on the Likert scale in Phase 1. The images of the needle injections are from a study by Lamm et al. (2010).

while, you will not be notified if it is correct or wrong what you click, no text on the screen. However, it will always be necessary to click it in the middle one before clicking the ones in the corners. Click start to begin the experiment.

The experimenter remained with the participants when they read the instructions, and if the participants had any questions, the part of the instruction relevant to their question was read aloud to them. No information other than what was stated in the instruction was provided. When the participants were ready to begin the experiment, they were told that the text "congratulations, you have now completed the experiment" would be displayed on the screen when they had finished this task, and they could go and get the experimenter. At the bottom of the on-screen instruction there was a grey box with the text "start." When the participants clicked the box, the

Fig. 3 Shows the bottles used in the choice situations as they were presented to the participants. The stimuli labeled on the bottles showed the B1, B2, and B3 stimuli used during conditional discrimination training and testing.

program initiated the conditional discrimination training. A sample stimulus appeared in the center of the computer screen. Clicking on the stimulus with the mouse cursor produced three comparison stimuli that appeared in three of the four corners of the screen. The sample stimulus remained in the center of the screen when the comparison stimuli were presented (simultaneous matching-to-sample). One corner of the screen always remained blank and the location of the blank corner was randomized for each trial. Clicking on one of the three comparison stimuli resulted in the immediate removal of the sample and the comparisons and the presentation of programmed consequences in the center of the screen. The programmed consequences consisted of written words. If the participants clicked on the stimulus defined as correct, the words "Excellent," "Very good," "Awesome," and "Well done" appeared. If the participant clicked on the stimulus defined as wrong, the word "wrong" appeared. The presentation of the programmed consequences lasted for 500ms followed by an intertrial interval (ITI) of 500ms. The conditional discriminations in Phase 2 were presented in an OTM training structure (AB/AC) with a simultaneous protocol and a concurrent presentation of baseline relations. Each conditional discrimination was presented five times, resulting in blocks of 30 training trials. The trained relations were **A1**/$\underline{B1}$-B2-B3, **A2**/B1-$\underline{B2}$-B3, **A3**/B1-B2-$\underline{B3}$, **A1**/$\underline{C1}$-C2-C3, **A2**/C1-$\underline{C2}$-C3, **A3**/C1-C2-$\underline{C3}$ (the sample stimulus is written in bold, whereas the defined correct comparison stimulus is underlined). The criterion was set to 90% correct responding in accordance with the experimenter-defined relations within one block. When the criterion was met with 100% probability of programmed consequences, the thinning of these consequences was set to steps of 75%, 50%, and 0% before the test for emerged relations was initiated.

Phase 3: Test for emergent relations When the participants responded with an accuracy of 90% or more correct in the last training block with 0% programmed consequences, the test for emergent relations was initiated. The test consisted of 90 trials including 30 trials of each of baseline, symmetry, and equivalence relations, respectively. The trials were presented in a randomized order throughout the test. The tested baseline trials were **A1**/$\underline{B1}$-B2-B3, **A2**/B1-$\underline{B2}$-B3, **A3**/B1-B2-$\underline{B3}$, **B1**/$\underline{C1}$-C2-C3, **B2**/C1-$\underline{C2}$-C3, **B3**/C1-C2-$\underline{C3}$, symmetry trials were **B1**/$\underline{A1}$-A2-A3, **B2**/A1-$\underline{A2}$-A3, **B3**/A1-A2-$\underline{A3}$, **C1**/$\underline{A1}$-A2-A3, **C2**/A1-$\underline{A2}$-A3, **C3**/A1-A2-$\underline{A3}$, and equivalence trials were **C1**/$\underline{B1}$-B2-B3, **C2**/B1-$\underline{B2}$-B3, **C3**/B1-B2-$\underline{B3}$, **B1**/$\underline{C1}$-C2-C3, **B2**/C1-$\underline{C2}$-C3, **B3**/C1-C2-$\underline{C3}$. The criterion was set to 90% correct responding in each of the relation types to respond in accordance with stimulus equivalence.

Phase 4: Class expansion In this phase, the tailored D stimuli (D1, D2, and D3) from Phase 1 were trained to the A stimuli

(A1, A2, and A3). Each relation was presented five times, resulting in training blocks of 15 trials per block. The stepwise thinning of the programmed consequences remained the same, and participants had to obtain a criterion of 93% or more correct for each relation type to proceed through the training blocks.

Phase 5: Test for emergent relations When the participants had completed the last training block with 0% chance of programmed consequences, a test for emergent relations was initiated, testing for the emergence of three 4-member equivalence classes. Each relation was mixed and tested five times in a 180-trial testing block. In addition to the tested relations in Phase 3, the tested baseline trials now included D1/$\underline{A1}$-A2-A3, D2/A1-$\underline{A2}$-A3, D3/A1-A2-$\underline{A3}$, symmetry relations A1/$\underline{D1}$-D2-D3, A2/D1-$\underline{D2}$-D3, A3/D1-D2-$\underline{D3}$, transitivity trials D1/$\underline{C1}$-C2-C3, D2/C1-$\underline{C2}$-C3, D3/C1-C2-$\underline{C3}$, D1/$\underline{B1}$-B2-B3, D2/B1-$\underline{B2}$-B3, D3/B1-B2-$\underline{B3}$, and equivalence trials C1/$\underline{D1}$-D2-D3, C2/D1-$\underline{D2}$-D3, C3/D1-D2-$\underline{D3}$, B1/$\underline{D1}$-D2-D3, B2/D1-$\underline{D2}$-D3, B3/D1-D2-$\underline{D3}$.

Phase 6: Choice test The choice test followed the same procedure as described in Arntzen et al. (2016a). Immediately after the participants had finished the test for emergent relations in Phase 5, the experimenter guided the participants to the cubicle located next to where they had been performing the conditional discrimination training and testing. The cubicle contained a table with three bottles of water. The participants were asked "to choose a bottle" and bring it to the experimenter who was waiting outside the cubicle. The three bottles were labeled with printouts of the three B stimuli (B1, B2, and B3). The participants choice was written down, before the participants were seated at a table to complete the last phase of the experiment.

Phase 7: Posttest evaluation of B stimuli In this phase, the participants were handed sheets with the B stimuli with similar Likert scales as in Phase 1 to rate the degrees of painfulness of the B stimuli after the participants had finished the conditional discrimination training and testing. They were instructed to "rate each of the three images on the corresponding scale." The experimenter remained out of sight of the participants while they rated the degree of painfulness of the images.

Statistical analyses Likert-type data are best treated as an ordinal scale. Because of the small sample and violation of homogeneity, a related-sample Wilcoxon Signed Rank test was run. This was to evaluate if the median ratings among the D1-B1 and D2-B2 stimuli were significantly similar. The expected degree of painfulness by rating for neutral stimuli would be 1 ("not painful"). The same tests were run to identify if the B-stimuli ratings were significantly different from each other. In addition, a Fisher's exact test was run to test for a significant

difference between number of choices towards bottle B1 and the sum of choices towards bottles B2 and B3.

Results

Trials to Criterion and Equivalence Class Formation

Fifteen participants formed three 3-member classes before successfully expanding the classes to three 4-member classes by training three D stimuli to the A stimuli. Number of training blocks varied across participants based on performance. The mean number of training trials required to establish the conditional discriminations (ABC) were 286 trials ($SD =$ 162.8). In the class expansion (Phase 4), the mean number of trials were 82 trials ($SD =$ 24.4). P13332 had the largest number of trials (150) in Phase 4. P13332's responding during Phase 4 was subjected to a trial-type response analysis. Before obtaining the mastery criterion, the participant responded in accordance to the experimenter-defined relations D1A1 and D2A2 in 21 and 24 out of 35 trials, respectively. For the D3A3 relation, the participant responded in accordance to experimenter-defined relations in 32 out of 35 trials.

Stimulus Ratings

The degree of painfulness ratings of the D and B stimuli are shown under the Likert Rating columns in Table 1. Eight out of 15 participants rated the D1-B1 and D2-B2 stimuli the same. That is, the D1 and B1 stimuli and the D2 and B2 stimuli were given the same value on the Likert scale. In addition, the eight participants rated stimulus B3 equivalent to the Q tip image with the value of 1 (not painful). In total, 13 out of 15 participants rated B3 as not painful. A Wilcoxon Signed Ranks test was run to test if there was is a significant difference in the median ratings between the D1-B1 and D2-B2 stimuli. The test showed that the median B1 ratings were not statistically significantly different from the median D1 ratings $Z = -1.82$, $p > .068$. The same was seen for the D2-B2 pairs $Z = -,33$ $p > .74$. The ratings of the B stimuli were significantly different: B1-B3 $Z = -3.25$ $p < .001$, B2-B3 $Z = -2.81$ $p < .005$.

Choice test Figure 4 shows an overview of the choices towards the bottles. Two participants chose the bottle labeled with the B1 stimulus (13%), seven chose the bottle labeled with the B2 stimulus (46%), and six participants chose the bottle labeled with B3 (40%) indicating that participants avoided choosing the bottle labeled with the B1 (most painful). Testing avoidance towards choosing bottle B1 was done with a Fisher's Exact test. We tested choices towards bottle B1 (2) versus the sum of choices towards bottle B2 and B3 (13), $OR = 0.02$, 95% CIs [0.00, 0.20], $p < .001$.

Discussion

The present study employed a perceived painful evaluation procedure that tailored the D1 and D2 stimuli (images of needle injections) to each participant. ToF was seen when the classes were successfully expanded and tested for. Most participants avoided choosing bottle B1 equivalent to the image rated as most painful. This way of testing for the transfer of avoidance in equivalence classes might be a valuable addition to previous findings (Augustson & Dougher, 1997; Garcia-Guerrero, Dickins, & Dickins, 2014).

Tailoring the Stimuli

Individually tailoring the stimuli for each participant might help reduce the stimulus control topography discrepancy between the arranged contingencies and the actual controlling properties. Eight of 15 participants rated the degree of painfulness for the D1-B1 stimuli and the D2-B2 stimuli the same even though the ratings were not significantly different. The remaining participants did not rate the above mentioned classes the same. There might be several reasons for this discrepancy. Most of the participants lowered their score for the B stimuli, except for P13332, P13221, and P13330. P13332 had the highest pain rating scores for both the D and the B stimuli. The participant also increased the pain rating scores from 5 (D1) and 4 (D2) to 5 (B1) and 5 (B2). During debriefing, the participant reported feeling very uncomfortable towards needles and needle injections and reported turning away from the screen during many of the trials in the class expansion phase (DA training). The median trials to criterion during the class expansion phase was 75, whereas P13332 had 150 trials to criterion during this phase. The participant turning away from the screen during the class expansion phase and not being able to attend to the controlling features of the stimuli could explain the high number of training trials to criterion for this participant. Other experiments have shown how attending behavior could be influenced by presenting aversive stimuli. For example, Dougher, Hamilton, Fink, and Harrington (2007) found that some participants were mildly startled when presented with a stimulus in the same equivalence class as a stimulus previously paired with a mild electric shock. One participant tried to remove the shock electrodes. Tyndall, Roche, and James (2009) trained conditional discriminations for one group of participants where training stimuli were paired with aversive images. For the other group, training stimuli were paired with neutral images. The participants who were provided stimuli paired with aversive images required more training and testing trials in formation of equivalence classes.

Twelve out of 15 participants rated the degree of painfulness for the B3 stimulus (in the same equivalence class as the Q tip) as not painful (1); two participants rated it as a little painful (2);

**Table 1.** Summary of Responding and Likert Ratings

P#	Trials	BSL	SYM	EQ	EXP	BSL	SYM	TR	EQ	Likert Rating					Bottle
										D1	B1	D2	B2	B3	
13333	210	30	30	27	75	45	45	30	60	3	3	2	2	1	B3
13214	150	30	30	30	75	45	45	30	60	3	3	1	1	1	B2
13216	270	30	30	30	75	45	45	29	59	3	3	2	2	1	B3
13219	690	29	30	30	105	45	45	30	60	2	2	1	1	1	B2
13226	180	29	29	30	60	45	45	30	60	4	4	2	2	1	B2
13223	240	30	30	30	75	45	45	30	60	3	3	1	1	1	B2
14130	210	30	30	29	60	45	45	30	60	3	3	2	2	1	B3
13229	210	30	30	30	75	45	45	30	60	3	3	2	2	1	B3
13213	660	30	30	27	120	45	44	29	59	3	2	3	3	3	B1
13332	390	30	30	29	150	45	45	29	59	5	5	4	5	2	B3
13221	240	28	30	30	60	45	45	30	60	5	4	1	3	1	B2
13336	150	30	30	30	60	45	45	30	60	3	1	2	1	1	B2
13225	180	30	30	30	75	45	45	30	60	4	2	3	2	1	B1
13330	240	30	30	30	75	45	45	30	60	3	4	1	2	2	B3
14127	270	30	30	27	90	45	45	28	60	5	3	3	2	1	B2

Note. Column P# represents participant numbers. Column Trials shows the number of trials to reach the first test for emergent relations in Phase 3. Columns BSL (baseline), SYM (symmetry), EQ (equivalence), and TR (transitivity) show the number of correct responses for each relation type of baseline, symmetry, equivalence and transitivity respectively. Maximum number of possible correct for BSL, SYM, and EQ in Phase 3 is 30. Maximum number of possible correct for BSL and SYM in Phase 5 test is 45. Maximum for TR is 30 and for EQ is 60. Column EXP shows the number of trials during class expansion in Phase 4. Columns under Likert Rating represents the Likert rating towards stimuli D1, D2, B1, B2, and B3, respectively (5 = most painful, 1 = least painful). Column Bottle shows the chosen bottle with the respective B-stimulus printout in the choice test (Phase 6)

and one participant rated it as moderately painful (3). Because we had no control group to evaluate the image of the Q tip or the B stimuli, it is not possible to know with certainty that the ratings are an indication of the B3 stimulus being equivalent to D3. It could be that the participants did not perceive stimulus B3 as painful. In contrast, when comparing the ratings of B3 to B1, and B2, it supports the notion that the B3 stimulus is equivalent to the Q tip, and not perceived as painful.

Choice Test

The choice test showed a significant avoidance for bottle B1 when the choices towards B2 and B3 are summed. Only 2 out
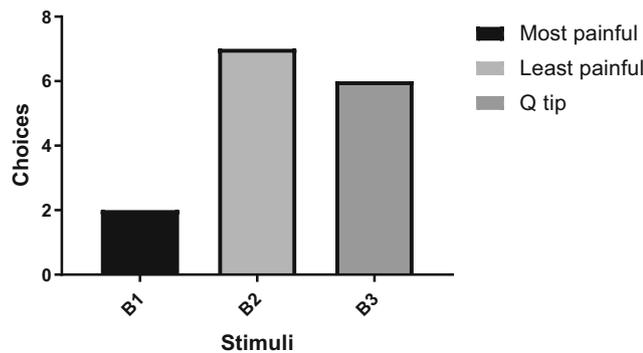
Fig. 4 Shows the number of participants who chose either B1, B2, or B3.

of 15 participants chose the bottle labeled with stimulus B1. P13213 chose bottle B1, but the participant also rated the degree of painfulness for stimulus B1 with the score of 2 (slightly painful), which was the lowest pain rating score of the three B stimuli for this participant. For the seven participants who chose bottle B2, four had rated the degree of painfulness for stimulus B2 as 1 (not painful). One interpretation could be that the 13 participants avoided choosing bottle B1 but chose either B2 or B3 because both stimuli were in classes related to stimuli (D2 and D3) with less degree of painfulness. However, this argument contrasts with the evaluation test where 12 participants rated the degree of painfulness for B3 as 1 (not painful) even if all the participants formed the three 4-member equivalence classes. In conclusion, one participant chose bottle B1, and three chose bottle B2 even though they rated bottle B3 as less painful. Similar results is not uncommon in ToF studies. For example, Amd and Roche (2017) used blurred facial stimuli to establish happier than relational series (X > A > B > C > D > E > Y) with the C stimuli functioning as the participants own blurred face. They found that despite participants learning that B is happier than C, they reported C as happier than B, indicating that participants learning history with stimulus C (their own face) override the intended experimenter-defined relational structure.

Previous studies from our lab have also shown similar results, for example, in Arntzen et al. (2016a), 20 participants in

Group 1 chose among bottles B1, B2, and B3. B1 was equivalent to a sunny weather symbol, whereas B2 was equivalent to cloudy, and B3 was equivalent to rain/thunder. Eleven participants chose B1, five chose B2, and four chose B3 indicating that the participants here treated bottle B2 and B3 as similar. Overtraining have been seen to increase the strength of equivalence classes (Bortoloti, Pimentel, & de Rose, 2014; Travis, Fields, & Arntzen, 2014) and also measured by ToF (Bortoloti, Rodrigues, Cortez, Pimentel, & De Rose, 2013). Thus, overtraining could increase the differences between B2 and B3 choices. These variations in the preference or choice tests might also be controlled for by presenting only two bottles (e.g., B1 and B3).

Participant Variability

Eight out of 15 participants rated the degree of painfulness for the D1-B1 and D2-B2 stimuli the same even though the ratings were not significantly different. Between-participant variability is commonly observed in ToF experiments (e.g., Barnes-Holmes et al., 2000; Dougher et al., 1994). For example, in Dougher et al. (1994) Experiment 1 only five out of eight participants showed a ToF measured by skin conductance, and six out of eight showed a ToF with regards to skin conductance level change. In addition, four of the participants showed noticeable skin conductance towards the stimuli in the second class (not equivalent to the stimulus pared with electric shock). Furthermore, in Experiment 1 in Barnes-Holmes et al. (2000), 16 out of 27 participants who formed equivalence classes rated the HOLIDAYS cola as more pleasant than the CANCER cola. However, 11 participants rated the CANCER cola more pleasant. Barnes-Holmes et al. discusses the differences in the learning history if the individual participants related to the words CANCER and HOLIDAYS.

Limitations and Further Research

The present experiment has some limitations. One limitation could be the similarity of the six valenced stimuli. Future studies should further investigate the tailoring of stimuli by adding a wider array of valenced stimuli the participant can rate. This might be time consuming, but it would further control for the variability in each participant's learning history.

A second limitation is that some additional measures are not included. For example, the participants could be asked to taste the water in the water bottles and consecutively rate the pleasantness of the water as in previous studies (Barnes-Holmes et al., 2000; Smeets & Barnes-Holmes, 2003) and more recently with different types of food (dos Santos & de Rose, 2018a). The next step would be to investigate whether the valence can be reversed or changed.

A third limitation might be that the choice test being presented before the B stimuli evaluation test influenced the ratings of the B stimuli. Future studies should counterbalance the order of the evaluation and choice tests. Finally, one argument could be that the choice test including water bottles is unsuitable with respect to the painfulness ratings. The bottles were used in the present study to keep as many variables as possible constant with respect to previous studies with similar arrangement. Further experiments could also manipulate the choice test as done in other studies (dos Santos & de Rose, 2018b).

Further research might also give some answers to how stimuli can evoke emotional responses such as fear, anxiety and avoidance towards stimuli the participants never have had any direct experience with (Friman et al., 1998; Lewon & Hayes, 2014). This is especially interesting when the relations among the stimuli are arbitrary, and the stimuli are nonperceptual (Dymond et al., 2015). Another needed experiment to do, is to employ the procedure described by Sidman et al. (1985) to train two stimulus sets with separate three 3-member classes and test for the emergence of one 6-member class where one or three members of one class could be tailored stimuli as in the current procedure. Preference tests could be admininstered as in the current experiment to assess if there would be any different outomes in three 3-member classes than in a merged three 6-member class.

Furthermore, future studies should also test how different training structures affect the ratings and choices. Studies have reported that the within class generalization of stimulus function decrements as a direct function of number of nodes (Bortoloti & de Rose, 2009; Fields, Hobbie-Reeve, Adams, & Reeve, 1999; Moss-Lourenco & Fields, 2011). Thus, training structure and class size should be systematically varied to observe how it affects ToF measurements. Another test for effect of number of nodes could be to present a within class preference test. This could be done by training and testing for three 4- or 5-member classes in a LS training structure with A1, A2, and A3 as stimuli of different tailored valence. Bottles could then be presented as B1, C1, and D1.

Conclusion and Implications

Based on the results in the present experiment, the stimulus evaluation procedure can be used to assess and choose stimuli that could be used in similar experiments by having the participants choose their own stimuli. This procedure could ensure that the valenced stimuli holds the connotative meaning that is experimentally intended.

One implication of the current study is highlighting the advantage of using individually tailored stimuli. Tailoring the painful stimuli for each participant could decrease the between-participant variability and might be important in small sample studies given interindividual variance in how

stimuli are perceived. In addition, the findings add to the existing body of research on using preference tests (e.g., Arntzen et al., 2016b; dos Santos & de Rose, 2018b), which is an important way to experimentally investigate how preferences can be altered and influenced towards stimuli in equivalence classes. Finally, the present procedure confirms an effective experimental arrangement to form and expand equivalence classes with meaningful or valenced stimuli. Therefore, these findings might help to shed some light on the environment behavior mechanisms causing the expansion of fear and avoidance behavior in anxiety disorders.

Compliance with Ethical Standards

Conflict of Interest The authors declare that there is no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards 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.

Availability of Data and Materials Data can be obtained upon request.

References

Amd, M., & Roche, B. (2017). Transforming valences through transitive inference: How are faces emotionally dissonant? *Quarterly Journal of Experimental Psychology, 70*, 2478–2496. https://doi.org/10.1080/17470218.2016.1246576.

Arntzen, E., Eilertsen, J. M., & Fagerstrøm, A. (2016a). Preferences in equivalence classes by low potency benign valenced stimuli. *European Journal of Behavior Analysis, 17*, 142–153. https://doi.org/10.1080/15021149.2016.1247637.

Arntzen, E., Fagerstrøm, A., & Foxall, G. R. (2016b). Equivalence classes and preferences in consumer choice. In G. R. Foxall (Ed.), *The Routledge companion to consumer behavior analysis* (pp. 65–77). New York, NY: Routledge.

Augustson, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy & Experimental Psychiatry, 28*, 181–191. https://doi.org/10.1016/s0005-7916(97)00008-6.

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*, 493–511. https://doi.org/10.1007/bf03395367.

Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record, 59*, 563–590. https://doi.org/10.1007/bf03395682.

Bortoloti, R., Rodrigues, N. C., Cortez, M. D., Pimentel, N., de Rose, J., & C. (2013). Overtraining increases the strenght of equivalence relations. *Psychology & Neuroscience, 6*, 357–364. https://doi.org/10.3922/j.psns.2013.3.13.

Bortoloti, R., Pimentel, N., de Rose, J., & C. (2014). Electrophysiological investigation of the functional overlap between semantic and equivalence relations. *Psychology & Neuroscience, 7*, 183–191. https://doi.org/10.3922/j.psns.2014.015.

dos Santos, S. L., & de Rose, J. C. (2018a). Investigating the impact of stimulus equivalence on children's food choice and preference. *Temas em Psicologia, 26*, 1–14. https://doi.org/10.9788/TP2018.1-01.

dos Santos, S. L., & de Rose, J. C. (2018b). Influence of cartoon characters on children's food preference via transfer of functions. *The Psychological Record, 69*, 153–163. https://doi.org/10.1007/s40732-018-0327-7.

Dougher, M. J., & Markham, M. R. (1996). Stimulus classes and the untrained acquisition of stimulus functions. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 137–152). Amsterdam, The Netherlands: Elsevier.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*, 331–351. https://doi.org/10.1901/jeab.1994.62-331.

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 Behavior, 88*, 179–197. https://doi.org/10.1901/jeab.2007.45-05.

Dougher, M. J., Twohig, M. P., & Madden, G. J. (2014). Editorial: Basic and translational research on stimulus-stimulus relations. *Journal of the Experimental Analysis of Behavior*, 1–9. https://doi.org/10.1002/jeab.69.

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

Dymond, S., Schlund, M. W., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: Symbolic generalization of human avoidance learning. *Behaviour Research & Therapy, 49*, 614–621. https://doi.org/10.1016/j.brat.2011.06.007.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. *Behavior Therapy, 46*, 561–582. https://doi.org/10.1016/j.beth.2014.10.001.

Eilertsen, J., & M., & Arntzen, E. (2017). Transfer of stimulus functions. *Norsk Tidsskrift for Atferdsanalyse, 44*, 33–49. Retrieved from www.nta.no.

Fields, L., & Garruto, M. (2009). Optimizing linked perceptual class formation and transfer of function. *Journal of the Experimental Analysis of Behavior, 91*, 225–251. https://doi.org/10.1901/jeab.2009.91-225.

Fields, L., Hobbie-Reeve, 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*, 703–723. https://doi.org/10.1007/bf03395336.

Friman, P. C., Hayes, S. C., Wilson, K., & G. (1998). Why behavior analysts should study emotion: The anxiety example. *Journal of Applied Behavior Analysis, 31*, 137–156. https://doi.org/10.1901/jaba.1998.31-137.

Garcia-Guerrero, S., Dickins, T. E., & Dickins, D. W. (2014). The gradual extinction of transferred avoidance stimulus functions. *Psychological Record, 64*, 581–599. https://doi.org/10.1007/s40732-014-0062-7.

Hayes, S. C., Devany, J. M., Kohlenberg, B. S., Brownstein, A. J., & Shelby, J. (1987). Stimulus equivalence and the symbolic control of behavior. *Mexican Journal of Behavior Analysis, 13*, 361–374 Retrieved from http://rmac-mx.org/wp-content/uploads/2013/05/Vol-13-N-3-361-374.pdf.

Lamm, C., Meltzoff, A. N., & Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 22*, 362–376. https://doi.org/10.1162/jocn.2009.21186.

Lewon, M., & Hayes, L. J. (2014). Toward an analysis of emotions as products of motivating operations. *The Psychological Record, 64*, 813–825. https://doi.org/10.1007/s40732-014-0046-7.

Markham, M. R., Dougher, M. J., & Augustson, E. M. (2002). Transfer of operant discrimination and respondent elicitation via emergent relations of compound stimuli. *The Psychological Record, 52*, 325–350. https://doi.org/10.1007/bf03395434.

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst, 26*, 195–213. https://doi.org/10.1007/bf03392076.

Moss-Lourenco, P., & Fields, L. (2011). Nodal structure and stimulus relatedness in equivalence classes: Post-class formation preference tests. *Journal of Experimental Analysis of Behavior, 95*, 343–368. https://doi.org/10.1901/jeab.2011.95-343.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. *Matching to sample: An expansion of the testing paradigm. Journal of the Experimental Analysis Behavior, 37*, 5–22. https://doi.org/10.1901/jeab.1982.37-5.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43*, 21–42. https://doi.org/10.1901/jeab.1985.43-21.

Smeets, P. M., & Barnes-Holmes, D. (2003). Children's emergent preferences for soft drinks: Stimulus-equivalence and transfer. *Journal of Economic Psychology, 24*, 603–618. https://doi.org/10.1016/S0167-4870(03)00004-7.

Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 102*, 47–65. https://doi.org/10.1002/jeab.91.

Tyndall, I. T., Roche, B., & James, J. E. (2009). The interfering effect of emotional stimulus functions on stimulus equivalence class formation: Implications for the understanding and treatment of anxiety. *European Journal of Behavior Analysis, 10*, 121–140. https://doi.org/10.1080/15021149.2009.11434320.