Running head: THE GRADUAL EXTINCTION OF TRANSFERRED AVOIDANCE STIMULUS FUNCTIONS

The Gradual Extinction of Transferred Avoidance Stimulus Functions

S. Garcia-Guerrero[a] Department of Psychology, Middlesex University, London

NW4 4BT: santiago.garciadefilippis@gmail.com[1]

T.E. Dickins[b] Department of Psychology, Middlesex University, London

NW4 4BT: t.dickins@mdx.ac.uk

D.W. Dickins[c] School of Psychology, University of Liverpool, Liverpool

L69 7ZA: dickins@liv.ac.uk

[1] Corresponding Author's permanent address: London SE21 8NH; 27, Pymers Mead. Phone: (+44) 02086702717

Abstract

We investigated the transfer of conditioned avoidance functions through equivalence relations, and the extinction of these functions, facilitated by verbal prompts. Nine participants acquired three 4-member stimulus equivalence classes using a matching-to-sample procedure. One class stimulus was paired, by classical conditioning, with an aversive tone, which was used in avoidance training of a distinct response. There were two groups: A established the equivalence classes before avoidance training, vice versa for B. During some avoidance trials, each stimulus presentation was followed by the request for a verbal estimation of the probability of the tone. The last trials, run in extinction, included a verbal prompt to corroborate the provided estimation. One participant in each group received no verbal prompts. To negate the necessary reliance on instructions-governed performance, an additional participant completed the experiment with minimal instructions. All participants who had the equivalence training prior to the conditioning showed within-class transfer of avoidance functions, in contrast to the others. All prompted participants who demonstrated transfer showed gradual response extinction, but with a differential gradient: responding decreased more sharply to the indirectly related stimuli than to the directly paired stimuli. The clinical implications are discussed.

Research in equivalence relations (Sidman, 1971, 1994, 2000, 2009) has generated considerable interest among behaviour analysts for it explains behaviours that lack a direct history of learning. Given initial conditional discrimination training with a particular set of stimuli, a participant then displays novel discriminative behaviour among the stimuli that were not directly trained. When such a phenomenon occurs, it is said that the stimulus set becomes a *stimulus equivalence class* (Sidman & Tailby, 1982). Namely, stimuli can functionally substitute for one another, including not only a stimulus with discriminative functions (S^D) but also the response and reinforcer (see Sidman, 2000). In other words, the entire contingency becomes equivalent.

A simplified way of representing the contingencies in an equivalence arrangement is as follows. Given stimulus *A* (as discriminative context) and stimuli *B* and *D* (as possible selection responses), the selection of *B* is followed by a specified reinforcer or consequence *C* (usually a kind of feedback):

$$(A \rightarrow B = C^b)$$

and given stimulus *A*, selection of *D* is followed by a specified consequence (C^d):

$$(A \rightarrow D = C^d)$$

As a result of this discriminative history, a participant displays the following three types of novel behaviour:

a) *reflexivity*: selection of *A* upon presentation of the same stimulus as context

$$(A \leftrightarrow A = C^x)$$

b) *symmetry*: reversal of the relations, so that given *B* selects *A*, or given *D* selects *A*

$$(B \rightarrow A = C^x) - (D \rightarrow A = C^x)$$

c) *transitivity*: selection of one stimulus in terms of another indirectly related, so given *B* selects *D*, and vice versa

$$(B \rightarrow D = \mathrm{C}^x) - (D \rightarrow B = C^x)$$

But as the variables specified experimentally are not the only ones exerting influence, control by unspecified stimuli (X) and consequences is acknowledged ($X \rightarrow C^x$). Thus the consequential conditions for all of the emergent responses (i.e., reflexivity, symmetry, and transitivity) are unspecified (C^x). This is due to the fact that these responses (i.e., selections) are produced without requiring explicit training and are tested in conditions of non-reinforcement; otherwise such discriminations would be attributed to a –scheduled– history of reinforcement.

Formation of a stimulus equivalence class enables members of that class to “share” stimulus properties (Goldiamond, 1962), functionally substituting for one another, as discriminative stimuli (S^D) or as eliciting stimuli. For example, using the same (*A*, *B*, and *D*) stimuli as before, if stimulus *B* then acquired additional functions (through discriminative training or contingent pairing) here referred to as ^{Y}, this would result not only in B^Y, but also in D^y and A^y, by virtue of equivalence class membership[2].

It is customary rigorously to satisfy the relational criteria (reflexivity, symmetry and transitivity) among the stimuli in equivalence studies (Sidman & Tailby, 1982; Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan, 1982; Sidman, Kirk, & Willson-Morris, 1985; Sidman, Wynne, Maguire, & Barnes, 1989). Fields, Adams, Verhave and Newman, (1993) however, have suggested that the transfer of stimulus functions is a more reliable

[2] The differences in the superscript notation between B as the stimulus acquiring the new stimulus properties directly, and D and A indirectly, is to illustrate the premise that despite the topographical similarities in the elicited responses, the “shared” stimulus properties will vary across some parameters. They are not considered identical; just as the properties of a conditioned stimulus (and response) are not identical to those of an unconditional stimulus (or reflex).

indicator of stimulus relatedness, as tests on emergent relations “can be passed even if the stimuli in a class are not all equally related to each other.” (p. 43).

Some scholars have addressed the transfer of eliciting functions within equivalence classes using both classical conditioning and conditional discrimination procedures. DeGrandpre and Bickel (1993) demonstrated transfer of functions associated with a nicotine reward across an equivalence class that contained the original discriminative stimuli. Dougher, Augustson, Markham, Greenway, and Wulfert, (1994, see also Augustson, Dougher & Markham, 2000), using extra-dermal activity measures, demonstrated transfer of a conditioned fear response across an equivalence class. Such transfers have also been demonstrated for avoidance responding (Augustson & Dougher, 1997), sexual arousal (Barnes & Roche, 1997) and mood (Barnes, Holmes, Smeets, & Luciano, 2004).

In these studies, all stimuli once established as members of a functional class, seem to be deemed *functionally equal* across dimensions; namely, to the extent that the elicited response is considered the same (cf. Fields & Moss, 2008). Differences in response to stimuli whose contingent relation is direct or derived might be of relevance in applied settings (e.g., Dymond & Roche, 2009) as well as basic research (e.g., Dymond & Rehfeldt, 2000). The following study addresses this issue. In particular, it is hypothesized that avoidance responses to the initial classically conditioned stimulus will withstand more extinction trials compared with indirectly related –functionally equivalent– stimuli. The following questions are also addressed: Is extinction of derived stimulus functions facilitated through verbal prompts? And, is there consistency between participants’ verbal estimations and avoidance responses?

Since the conception of aversive conditioning is usually equated to painful stimulation, considering the nature of the US employed in the present study, the term *conditioned avoidance response* is adopted (i.e., the key presses turn off the tone). Therefore, when referring to the responses to indirectly related stimuli (i.e., key presses to stimulus C1 and D1), the term *derived conditioned avoidance responses*[3] will be used.

Method

Participants

Fifteen human participants, aged between 26 and 47 (age M = 31; SD = 5.16), were recruited through personal invitation. Due to the strict training criteria, six participants were removed from the study. Eight of the remaining participants were equally assigned to arrangement A (participants 1, 2, 3 and 11) and B (participants 7, 8, 9 and 12). Each arrangement presented the experimental phases in a different order. Two of these participants (11 and 12) served as point of comparison for the presence of the verbal prompts: they received an extended phase 5 instead of phase 6 (see details below). Finally, the ninth participant (15) was exposed to the experimental arrangement A but without the instructions about the tasks (i.e., only given the minimum guidance necessary to operate the software application –see details below).

Each participant was asked to read and sign a consent form containing information about exposure to a distressing but innocuous high pitched tone during some phases of the experiment, as well as the other aspects of the tasks. The experimental procedures were passed by the University of East London's School of Psychology Ethics Committee and deemed to conform to the codes of practice recommended by the British Psychological Society.

[3] By this, we do not mean that *responses* per se are derived. It is the –avoidance– discriminative stimulus property that is being derived. But for convenience, here we refer to the responses as a way of talking about them differentially. It is from observing these, after all, that we infer the derivation of such functional properties.

Apparatus and stimuli

A personal computer (windows-7 operating system) was used to run the software application specifically designed for the study. The stimulus presentation and data collection were programmed in *Visual Basic 6®*. Visual stimuli were displayed on a 15.6” monitor. The on-screen positions of the visual stimulus keys were the centre (sample) and each of the corners (comparison stimuli) of a central (15 x 15 cm) square (with 1 cm between the sample and the comparisons). When displayed, the comparisons were 9.5 cm away from the monitor’s lateral edge; when no comparisons were displayed the respective space remained blank. The tally of points gained was located at the bottom-centre of the screen, between the two lower-corner keys. Visual samples always appeared in the central position, and comparisons on the outer.

Visual stimuli consisted of 19 images mostly corresponding to mixed Cyrillic and Arabic letters (4^2 cm when displayed on the screen) arbitrarily grouped into four classes (1, 2, 3, 4). Each contained four class-members (A, B, C, D), plus three additional neutral stimuli (N5, N6, N7) used exclusively for *reflexivity* tests (Figure 1). As customary, unbeknownst to the participant, the stimuli were coded alphanumerically with letters referring to each stimulus member, and numbers denoting the stimuli classes.

INSERT FIGURE 1 ABOUT HERE

The aversive auditory stimulus consisted of a 5-second [91 dB (12000 Hz)] tone: the decibel levels could be reduced if necessary –none of the participants requested this. The tone was transmitted through Sony® stereo headphones (MDR-XD100) tested using a sound level meter CEL Instruments Ltd (model CEL-269) where a headphone coupler was required. Each participant wore the headphones throughout the experiment, including during matching-to-sample tasks where no tone was contingent, to isolate noise and to control for discriminative functions over the tone contingency at different phases of the experiment.

The task instructions were displayed on the screen (Arial 12 font) and there was also a hard copy of the instructions available beside the computer for necessary revisions during the tasks (except for participant 15 who only received minimal instructions).

Procedure

Each participant was verbally informed in general terms about what s/he would be doing without revealing the purpose of the study or specifics that would compromise the performance. Each was assured about the anonymous nature of their participation and a full explanation followed the termination of the experiment. Participants were offered an incentive: two entries for art exhibitions at the Victoria & Albert Museum, London (worth £25) for their participation. Additionally, a prize (worth £30) was offered at the end for the best performance in the tasks, consisting of two entries for any IMAX film at the Science Museum of London.

Participants received the order of the experimental phases in accordance with either arrangement A or B; the latter included phase 1 after phase 3 (see Figure 2). During each performance the experimenter quietly remained 3 m behind the participant, and out of sight to prevent the inadvertent cueing of responses. It was explained that interactions with the experimenter were not endorsed: his role was to intervene only when task setting was necessary and to answer questions regarding on-screen instructions. Occasionally, when a participant reported no understanding of what was asked of him/her, the experimenter would suggest reading the printed instructions, or he would rephrase them. Participants were asked for their permission to video-record the session and it was explained that only the experimenter would have access to the footage, which would be destroyed after data analysis was complete.

INSERT FIGURE 2 ABOUT HERE

The procedural description below follows the experimental arrangement A.

Phase 1: Stimulus Equivalence Training and Testing. Using a one-to-many (or sample-as-node) training structure (Arntzen, Grondahl, & Eilifsen, 2010) for the interrelated conditional discrimination procedure (Sidman, 1986, 1987, 2009), the baseline conditional relations for the emergence of three 4-member stimulus equivalence classes were trained (i.e., the A stimuli served as the node in each class, and the trained relations were A1-B1, A1-C1, and A1-D1; similarly for A2 and A3). The stimuli in class 4 served solely as incorrect comparisons for and thus remained untested. Their inclusion was to increase the likelihood of the intended discriminative stimulus control (i.e., not by exclusion) and avert false positives (Sidman, 1987; see also Sidman 2009; Carrigan & Sidman, 1992).

Once facing the computer monitor, the participant was given the following on-screen instructions (the excerpts within square brackets, in the instructions below and subsequent phases, constituted the minimal instructions for participant 15):

> [In a moment some figures will appear on the screen. Look at the image in the centre of the screen, click on it in order to make appear four ‘outer images’ in each of the corners of the screen. Select one of the four outer images by

clicking on it.] At the beginning the computer will give you feedback on every choice, and at other times it will not, but there is always a correct selection. Besides, you can make a correct selection in all the tasks without feedback by carefully attending to the tasks that come with feedback.

Even though the first tasks are easy, it is important to pay close attention as these will increase in difficulty, and choosing the correct figures in the latter part of the experiment will depend on the knowledge you gain during the early parts of the experiment.

[Your objective is to make as many correct selections as possible. If you have any question, please ask the experimenter now and when ready: Press the button below to continue.]

Nine relations (three AB, AC, and AD) were trained as shown in Table 1 (below). For instance, upon the presentation of A1, only the selection of B1 (and not of B2, B3 or B4) was reinforced in a particular baseline trial. All other choices –incorrect selections– were followed by information indicating an incorrect response –punishing feedback. Likewise, during tests of equivalence, in the presence of (say) D3 as sample, the correct selections would be B3 or C3, though here no information followed a response.

INSERT TABLE 1 ABOUT HERE

For each trial a *sample* stimulus appeared at the centre of the computer screen; the participant had to click on it (viz., an “observing response”) in order to make four choice stimuli or *comparisons* appear in the corners, with the sample remaining present. The participant then chose one of the comparisons by clicking on it. After this, the screen cleared immediately and a conditioned reinforcement –or punishment– was made contingent in the form of a 4^2 cm green tick image for correct selections and a red cross for incorrect ones. These feedback stimuli appeared in the centre of the lower half of the screen for 1.5 seconds, together with differential audible chimes for each of these. For every correct selection one point was added to the tally located at the bottom-centre of the screen. One point was subtracted for every incorrect selection. After a 1.5-second intertrial interval, the next set of stimuli appeared.

Given some of the artifactual effects of delayed-cue procedures traditionally employed (e.g. Glat, Gould, Stoddard, & Sidman, 1994), in the present study conditional discriminations were aided by a *correctional feedback* feature: initially, on the baseline trials, if an incorrect comparison was selected, all the incorrect comparisons disappeared, leaving behind the sample and the correct comparison for two seconds. This post-response cuing was only programmed to occur once for each sample, after the first incorrect response, whether or not this response had been preceded by correct selections.

In both training and testing, comparison arrays always comprised stimuli with the same alphabetic designation. Selection by exclusion was prevented by consistently using four comparisons per trial (Carrigan & Sidman, 1992; Fields, Verhave, & Fath, 1984; Sidman, 1987).

Each presentation of a sample and its comparisons constituted a trial type. Baseline trials (see Table 1) consisted of three trial-blocks (blocks AB, AC and AD, each consisting of three trial types). Throughout both training and testing the order of trials was randomized within blocks. In addition, trials were randomized across blocks with the constraint that all possible trial types had to occur before any were repeated, for example during mixed symmetry and transitivity probes. The screen position of the comparison stimuli was also randomized across trials to control any systematic relation between a comparison stimulus and its screen position (Sidman, 1987, 1992, 2009; cf. Sidman, 1992a), namely, the correct stimulus could not appear more than twice consecutively (sometimes thrice) in the same position. This constraint was to prevent a pattern of "elimination" by the participant or the possibility of following a rule on the lines of: "the correct selection in the next trial must be among the key positions thus far unselected". These parameters applied to all blocks of trials in both training and testing throughout the experiment.

For the baseline conditional discriminations, the learning criterion consisted of a minimum of five consecutive correct selections within each trial-block (i.e., a minimum of 15 consecutive correct selections in total to move on from a baseline trial-block). Training continued until participants reached the performance criterion twice, so a 15-trial baseline block was repeated once (in total 30 correct selections for all of the baseline trial-block, 10 per each single AB, AC and AD block relation). Then a 27-trial mixed-baseline block was presented where only one error was permitted and with reinforcement withdrawal ratio of .30; as preparation for the probes to be run in extinction. More than two incorrect selections led to a repetition, and if the participant failed this second attempt s/he was re-exposed to the previous unmixed blocks (see Figure 3).

INSERT FIGURE 3 [80% page size] ABOUT HERE

Once the baseline-blocks (AB, AC and AD) had been accomplished successfully, a series of unreinforced test probes was given. First a 36- trial block of *mixed* symmetry probes followed, with a learning criterion of only one incorrect selection permitted (35 correct out of 36), and repeated once if unsuccessfully completed. Failure to complete the second exposure to these tests entailed rehearsal of baseline relations.

Subsequently, a 72-trial block of *mixed* transitivity probes was presented, and repeated once if not successfully completed, which initially (for the first seven participants) meant attaining the criterion of 71/ 72 trials correct. From participant 8 onwards, owing to the difficulty of achieving such a high performance, this criterion was relaxed to 70/72, as long as the errors were not both within the same block.

Next, a 27-trial block of *mixed* symmetry and transitivity relations (one per each single relation) was presented, delivered for a second time if there was more than one error. If a participant failed to pass any probe block, s/he was re-exposed to baseline training and repeat testing. For repeat baseline training at this stage though, the consecutive learning criteria were halved (see Figure 3).

In view of the stringent learning criteria, participants re-exposed three times to baseline conditional discrimination trials, for not achieving the criteria, were rewarded for their participation and withdrawn from the study –additional rehearsals were thought likely to be conducive to intervening factors such as fatigue.

Finally, in order to satisfy the equivalence relations test criteria (Sidman & Tailby, 1982), an extra set of neutral stimuli was introduced (i. e., N5, N6, N7) to test reflexivity relations[4] (see Table 1). Reflexivity probes were run once per stimulus relation (a total of 12

[4] In the conceptual elaborations –and methodological notes– surrounding reflexivity (e.g., Sidman & Tailby, 1982 p. 6; Sidman, 1994 p. 130, pp. 168-169, p. 319-320) there is no allusion to the available comparison stimuli as a relevant feature (the definitive feature being that the stimulus "exhibits the same conditional relational with respect to itself", as long as it participated in the baseline conditional discriminations). Whilst it could be argued that utilizing novel stimuli as comparisons for reflexivity blocks can lead to selection by exclusion, such exclusion can also be made on the basis of the different histories of discriminative reinforcement for the stimuli used in the other relational blocks (besides, it is unknown whether the baseline "discriminate-able" contingencies could conflict with identity matching). Given some of these inconsistencies and multiple

identity matching trials), where the performance criterion was no more than one incorrect selection; however failure to achieve this did not lead to baseline rehearsals.

Phase 2: Aversive Classical Conditioning. A delayed classical conditioning procedure was implemented to establish the relation between a stress-eliciting tone (US) and stimulus B1 (as CS). The B1 (CS+) and B2 (CS-) stimuli were presented alone in the centre of the screen (for 10 seconds) four times each in a random sequence, with an interim blank screen varying between 2-8 seconds, in order to minimize temporal conditioning effects. A 5-second high pitch tone [91 dB (12000 Hz) at its peak] was used to elicit a presumptive stress response (UR). A total of four B1 and B2 presentations were programmed. The stimuli appeared in the centre of the screen as described for the matching-to-sample tasks.

The displayed instructions were [minimal instructions]:

> [In this phase you do not need to select any image. All you need to do is to pay attention to the screen until further instructions are given.]
>
> Some figures will appear on the screen, one at the time. It is important that you watch the figures carefully. [At times a sound may be played.]
>
> If you have any question, please ask the experimenter now, and when ready:
> Press the button below to continue.

Phase 3: Avoidance Training. Using a *delayed differential aversive conditioning procedure,* the participant received minimal instructions indicating that avoidance was possible based on responses made on the spacebar. The respective instructions were [minimal instructions]:

---

explanations around reflexivity (cf. Fields et al., 1993; Hayes, 1991; Steele & Hayes, 1991) we decided to explore such variables by modifying this methodological aspect. The respective data are not included in this paper.

> [As previously, in a moment some figures will appear and some will be followed by the tone you experienced before. However, this time, you can prevent the tone from playing by pressing the spacebar-key of your keyboard several times as soon as the image appears, when you think it is necessary.] If no key-presses are made during the first seconds, the tone will follow.
>
> It is important that you pay attention and concentrate on the screen at all times. If you have any questions, please ask the experimenter now. And when ready: Press the button below to continue.

Avoidance training consisted of a minimum of four presentations each of B1 and B2. The conditioning parameters were similar to the previous phase, except that a minimum of eight responses on the space-bar (FR-8) deactivated the tone, and the visual stimulus on the screen simply disappeared after the 10 seconds had elapsed. Fewer than eight responses, during the first five seconds of B1's presentation, resulted in exposure to the tone for the remaining 5 s. In order to prevent instruction-governed behaviour, the participant was not informed about the necessary number of spacebar presses to cancel the tone (i.e., FR-8). No stressful tone was contingent upon B2, regardless of the participant's responses. These "avoidance trials" were presented until the participant made four consecutive avoidance responses successfully (i.e., eight or more spacebar presses); presses during B2 however, counted as errors, thus re-setting the avoidance criterion. Consequently, the total number of exposures to the tone during this avoidance training could differ across participants. If a participant was exposed to the tone four times without demonstrating effective avoidance, the participant was prompted to review the instructions carefully and returned to the previous phase.

The presentation of class 3 stimuli (alongside the other randomised stimuli) during experimental phases 2 and 3 was thought to further prevent temporal conditioning effects and predictability during the respective trial-blocks.

Phase 4: Test for Transfer of Avoidance. Conserving the same parameters, transfer of avoidance function was tested (i.e., tone exposure if failure to fulfil the avoidance FR-8 criterion for B1 only). With the exception of the A stimuli (since it could be argued that any acquisition of “stressful” properties by the A stimuli could be attributed to sensory preconditioning or second-order conditioning effects), all of the stimuli from classes 1, 2 and 3 in equivalence training were presented in two trial-blocks. Stimulus presentations were quasi-random within each trial block (i.e., all stimuli must have appeared before any could repeat). The instructions were [minimal instructions]:

> As before, in a moment some figures will appear and some will be followed by the tone you experienced before. This time, however, with more figures involved.
>
> [Continue to press the space-bar several times, as soon as the image appears, to prevent the tone when you think it is necessary.] Press the button below to continue with the task.

Phase 5: Post hoc estimation of the probability of tone presentation. Each stimulus (classes 1, 2 and 3) was presented individually and upon each presentation the participants were asked to estimate the probability of the tone “happening” if s/he had not pressed the space-bar (in accordance with avoidance trails). The participant did this by clicking on the “estimation keys” (i.e., definitely or probably happening / definitely or probably not happening). Once a selection had been made, the next trial appeared after a 2-second intertrial period. This block was run once for all the stimuli (i.e, 12 stimulus presentations). The displayed instructions were as follows [minimal instructions]:

> During this phase, [continue to perform as you have been in accordance to the previous tasks. In addition, you will be asked the likelihood of the tone sounding.] Please follow the on-screen instructions. Press the button below to continue.

And after each stimulus presentation the following box appeared:

<table>
<tr><td colspan="4">How probable do you think the tone would have happened if no key-presses had been made?

Please rate by clicking on the options provided below</td></tr>
<tr><th colspan="2">Happening</th><th colspan="2">Not Happening</th></tr>
<tr><th>Definitely</th><th>Probably</th><th>Definitely</th><th>Probably</th></tr>
</table>

Phase 6: Post hoc probability estimation with feedback. Once participants had provided a probability estimation for each of the stimuli (1-3: A-D), trials of conditioned discriminated avoidance were presented in extinction. That is, B1 was no longer followed by the tone if no response occurred within the prescribed time period (i.e., 5 s); other than this, the parameters were the same as in phase 5.

This time, participants rated the probability of the tone presentation with only two scale values (i.e., would have happened / would not have happened). Once the participant had given an estimation, an additional window appeared wherein a challenge question – “Are you sure of your prediction? Would you like to corroborate it?” – offered the participant the possibility of corroborating her/his estimations about the likelihood of the contingent tone (viz., verbal prompts).

Participants could decide to corroborate their estimation. If they opted for this, they were either given three tallied points for a “correct” estimation, or they had three points deducted for an incorrect estimation. Further feedback was provided immediately in the form of a 4 cm^2 green tick or red cross image displayed below the answer buttons. Participants who decided not to corroborate their estimations still gained or lost points accordingly, but no further feedback was provided. The instructions and rating request were as follows [minimal instructions]:

> [Similar to the previous task some images might be followed by the tone unless you press the space-bar several times.] Again, we will ask your estimation about the probability of the tone happening, this time you will only have two choices as to whether you think the tone would have happened or not.
>
> However, an additional option is at play: you will be challenged and given the opportunity of corroborating your estimation before winning or losing points. You are given two options: 1. To go ahead and corroborate your estimation or 2., opting for not corroborating. [If you decide to corroborate your estimation, for each “correct” estimation you will win three points, and for each “incorrect” estimation you will lose three points.] The addition or subtraction of points will be immediate and feedback would be provided.
>
> [If on the other hand, you decide not to corroborate your estimation, you will win or lose one point for “correct” or “incorrect” estimations respectively.] In this case, [no feedback will be given.] the total amount of earned or lost points will be displayed at the end of the entire task; after several trials.

[Your aim is to make as many correct estimations as possible.] If you have any questions, please ask the experimenter now. And when ready: Press the button below to continue.

Rating request:

How probable do you think the tone would have happened if no key-presses had been made? Please rate by clicking on the options provided below	
HAPPENED	**NOT HAPPENED**

Whereupon the challenge question and corroboration option were as follows:

Are you sure of your prediction? Would you like to corroborate it?	
YES	**NO**
✓ or **X**	

This phase consisted of three blocks sufficient to reveal a gradual change in response pattern. Phase 6 was omitted for comparison participants, instead they continued to respond to an extended phase 5 which contained as many additional trials (i.e., 4 blocks in total), the last three blocks of which were presented in extinction.

Phase 7: Post-experiment equivalence tests. A final matching-to-sample trial block of mixed symmetry and transitivity probes was introduced to corroborate that the equivalence relations had endured throughout the experiment. The performance criterion was identical to the one employed in phase 1.

Results

What follows is an idiographic analysis. Bearing in mind that demonstrating equivalence was a prerequisite, Table 2 shows the accumulated number of errors over the total number of trials each participant had per trial-block type during the equivalence training.

INSERT TABLE 2 [and Table 2 continuation] ABOUT HERE

All participants completed the experiment in one session. The time of completion of the entire experiment ranged between 50 min and 2.5 hours.

Due to the programmed parameters of the avoidance training (phase 3), the number of exposures to the US (i.e., tone) varied across the participants. These were as follows for participant number and US exposures respectively: 1:2 (i.e., participant 1 had 2 exposures); 2:0; 3:0; 7:9; 8:0; 9:11; 11:3; 12:5; and 15:4. As participants were uninformed about the FR-8 necessary to deactivate the tone, these differences reflect sensitivity to scheduled contingencies.

Table 3 contains participants' data from arrangement A for both transfer tests and subsequent trials registering verbal estimations about the tone contingency, and also the corroboration trials run in extinction for stimulus classes 1 and 2. Stimuli A were excluded due to the possibility of either sensory preconditioning (arrangement A) or higher order conditioning (arrangement B) effects; explaining responses which would otherwise suggest transfer of stimulus functions across classes. Despite the fact that class 3 stimuli were also presented during all the trial-blocks represented in the results tables, these are not included as they received the same pattern of responding to class 2 stimuli –i.e., zero key-presses even though there was no differential discrimination training for them. The order of the stimuli within the tables is to facilitate its reading by contrasting stimuli against each class (i.e., 1 and 2). By no means is it meant to mirror the actual order in which each participant saw the respective stimuli, for they were randomized (see Appendix A for the actual order of stimuli presentation across conditions and participants).

In Table 3 (as well as in subsequent Tables 4 and 5), avoidance response constitutes the number of times each participant (i.e., 1, 2, 3, and 11) in arrangement A pressed the spacebar before each stimulus, represented by the first numeric values in the columns. The bracketed numbers (1-4) in the “estimation” column (third trial) correspond to the estimation-button pressed and its respective semantic value (i.e., 1 = “definitely happening”, 2 = “probably happening”, 3 = “definitely not happening” and 4 = “probably not happening”). The bracketed letter during the corroboration trials represents the dual value button chosen by the participant (i.e., Y stands for yes as in “happened”, and N for “not happened”). Finally, the plus + or minus – arithmetic symbols represent a positive or negative corroboration respectively with regard to the provided estimation. Therefore, an estimation of “happened” would produce a negative corroboration indicating that the provided estimation was wrong, given that corroboration trials were run in extinction. The participant not opting to corroborate is represented by a dot •.

INSERT TABLE 3 ABOUT HERE

Differential responding to members of class 1 and class 2 can be observed for all participants in arrangement A. Class 1 stimuli led to numerous spacebar presses, whereas class 2 stimuli received no key presses at all, as denoted by the numbers outside the parenthesis.

Interestingly, whilst estimating the tone contingency, participant 1 gave the indirectly related stimuli (i.e., C1 and D1) the same estimation value as B1 by pressing the button (1) which corresponded to “definitely happening”. All class 2 stimuli were given a (3) “definitely not happening” estimation. During corroboration trials however, the first negative corroboration to D1 seems to have led to a change in her subsequent “negative” estimation (N) to the next stimulus C1, yet still “avoiding” the “presumed tone” as manifested by the key presses (i.e., still demonstrating derived discriminative responding). The tone estimation for B1 does not seem to have been affected by this as the participant both avoided and received a negative corroboration for her estimation. From the fifth trial-block onwards, participant 1 did not make avoidance responses in the presence of any of the indirectly related stimuli and continued obtaining positive corroborations throughout, though she still avoided and got a negative corroboration for B1. By the sixth trial block, even responses to B1 were extinguished and a positive corroboration was received for her “negative” estimation (i.e., “not happened”). The high numbers were due to the key being held down instead of generating single presses.

Participant 2, whilst avoiding all three stimuli, she only assigned the estimation value (1) to B1, and (2) to C1 and D1. All class 2 stimuli were given an estimation value of (3). During the fourth trial-block negative corroborations upon C1 and D1 did not lead to a change in response to B1. In the fifth trial-block however, an effect can be noted only for the indirectly related stimuli where neither avoidance occurred nor negative corroborations followed. By the last trial-block complete response extinction is evident and the participant’s estimations are in accordance with the contingencies.

Participant 3 gave both the directly and indirectly related stimuli the same “probable” estimation value (2), in contrast to stimuli from class 2 whose estimation was (3). The estimation given to all class 1 stimuli suggesting the presence of the tone was constant during the fourth trial-block, thus obtaining a negative corroboration for each. However, this appears to have produced a drop in response to all stimuli for the remaining trials, for which the participant always received positive corroborations over his subsequent “negative” estimations.

Participant 11 served as a comparison for the foregoing participants from arrangement A, and the role of verbal prompts in facilitating contact with the contingencies in extinction; thereby behavioural change. Instead of receiving the last three corroboration trial-blocks, she had three additional estimation trial-blocks, except for these being run in extinction in which the tone contingency was not operating. From the beginning, a differential degree can be noted in her estimations where B1 and B2 were given the “definitely” value for each end, and the others received the “probable” value for the likelihood of the tone contingency. Avoidance responses were made throughout the extinction trials, except in the fifth trial-block before the D1 stimulus. As expected, the participant continued to respond during the trial-blocks run in extinction, where some response variability can be noted in the fifth trial-block (i.e., no responses made before D1 and a change in the estimation to D2). By the sixth trial-block, changes in the estimation given to class 1 are of relevance, all of which received a “probably happening” estimation value (2), in contrast to class 2 which received a “definitely not happening” estimation value (3).

In general, the responses of arrangement A participants to class 1 stimuli decreased from trial-block 5 for the indirectly related stimuli. By trial-block 6 all responses underwent extinction, including those to stimulus B1. This drop (i.e., zero presses) is correlated to the accuracy of the estimations given to each stimulus (i.e., reporting the “extinct” tone-contingency, as in “Not happened”) with one exception (i.e., participant 1, trial 5).

Table 4 shows the data for participants who received experimental arrangement B. To prevent overexposure to the tone –and possible habituation– during discriminated avoidance training, it was necessary for participants 7, 9 and 12, after failing to reach the avoidance criterion, to be returned to the classical conditioning phase and asked to re-read the instructions carefully (the participants seemed not to come under the control of the FR-8 schedule requirement). This recycling was done twice for participant 7 and once for participants 9 and 12.

INSERT TABLE 4 ABOUT HERE

Save for participant 7, none of the participants from this arrangement demonstrated the transfer of avoidance across the other members in class 1. Additionally, not only did no class 2 stimuli elicit key responses, but also the estimation given to them was that of “definitely not happening” (3) in a consistent manner.

Participant 7 was the only one from group B who demonstrated transfer of stimulus functions across equivalent classes. Unlike the other participants, she decided not to corroborate her estimations as trials progressed. In the first trial-block testing transfer, the participant did not respond to C1, presumably due to a brief distraction (as stated verbally upon questioning her at the end of the experiment). This did not happen before stimulus B1 and D1. In the fourth trial-block, D1 was the first stimulus followed by a negative corroboration. This seems to have led only to a change in the estimation given to C1.

However, B1 was avoided and followed by a negative corroboration. Immediately after the first positive corroboration to C2 the participant ceased corroborating for all class 2 stimuli; a pattern that stayed constant during the remaining trial-blocks. By the fifth trial-block, B1 was the only stimulus which still led to both key presses and negative corroboration. During this trial-block, corroboration dropped for the rest of the stimuli. By the sixth trial-block thorough response extinction was evident.

During the first trial-block testing transfer, participant 8 did not even respond to B1. During the second trial-block, she only effectuated presses before B1. In the third trial-block, the avoidance response registered as key presses is very weak, with only one key-press being recorded. In addition, the estimation given to class 1 varied in contrast to class 2 stimuli whose estimation value was (3). In the fourth trial-block, B1 is the only stimulus that receives key-presses, now in large numbers, followed by negative corroboration. All the other stimuli neither elicited key-presses nor did they get negative corroborations. The fifth trial-block shows a similar pattern with a change in estimation, followed by positive corroborations. By the sixth trial-block, no response is emitted and all corroborations produced correct estimations during extinction.

Participant 9 performed in a similar fashion to other participants during the first three trial-blocks, and assigned a positive estimation of “probable” to B1 and a “definite” negative estimation for all other stimuli. Unexpectedly, by the fourth trial-block, no key responses occurred; this seems responsible for immediate response extinction for the forthcoming trial-blocks. A positive estimation, followed by its corresponding negative corroboration, occurred for B1 during this and the next trial-block. Such response estimation changed by the sixth trial-block. When the participant was questioned on her lack of response upon B1’s presentation during the fourth trial-block, she said: “Well, I let it pass and that was it. I realized” (It is reasonable to assume then, that she accidentally contacted the null contingency between the visual stimulus and the tone, which led to an instantaneous extinction of the key-pressing response –in contrast to her estimations).

Comparison participant 12 responded to arrangement B without verbal prompts facilitating contingency sensitivity. Similarly to participants 8 and 9, B1 was the only stimulus upon which key-presses were emitted, including extinction trials. The estimation value (1) given to B1 is stable across all trial-blocks. Conversely, the initial estimation value given to the other stimuli C1 and D1 in the third trial-block, shifted from (4) to (2) for both stimuli (C1 and D1), and back to (4) by the sixth trail-block. The estimation value (3) given to class 2 stimuli was constant throughout all the trial-blocks.

Table 5 shows data for participant 15 who underwent the experimental arrangement A but with fewer instructions.

INSERT TABLE 5 ABOUT HERE

This participant behaved in an interesting way with respect to the corroboration response. The punishing effect from the negative corroborations led to a drop in such a response rather than to a change in either the key-pressing or the estimations that the corroborations were referring to; responses which remained stable. The participants' verbal report after completing the experiment confirms that (similarly to one of our pilot participants), his responding was controlled by the contingencies surrounding acquisition or loss of points: "I realized I was losing points every time my corroborations were wrong, so I stopped corroborating" –demonstrating sensitivity to contingencies other than those intended by the experimenter.

All participants successfully completed post-experiment matching-to-sample block-tasks containing mixed symmetry and transitivity probes (vide experimental phase 7).

Discussion

The data from participants in arrangement A corroborate those reported by Dougher et al. (1994), as well as Augustson and Dougher (1997), concerning the transfer of emotional responses acquired through classical conditioning (viz., transfer of avoidance evocation) across stimulus equivalence classes.

The results of this study differ from those reported in Roche and Barnes (1997) who found no effect of the temporal order in which conditioning and equivalence training took place upon the transfer of stimulus functions. The dependent variable they used to assess the transfer of functions was, however, phasic changes in skin resistance, instead of distinctive *overt* behavioural responses as implemented here. This may represent a critical difference in comparison with the present study in which all but one of the participants from arrangement B, who received respondent conditioning prior to equivalence training, failed to demonstrate the transfer of *avoidance* stimulus functions.

The present study contributes empirical data about the decline in the strength of *derived conditioned avoidance responses* during extinction (cf. Skinner, 1938/1991) within the framework of equivalence relations. In support of the hypothesis, the majority of participants in this study showed that the responses to stimulus B1, which had been directly paired with the aversive stimulation, were more resistant to extinction (i.e, more invalidating feedback was necessary before behavioural change was evident) than responses to indirectly related stimuli. For example, participant 1, 2, 7, and 8 did not drop their key-pressing response until receiving contingent negative feedback for the second time. Participant 3 did so after the first negative corroboration, showing greater sensitivity. Participant 9, as previously mentioned, fortuitously contacted the extinction contingency from the moment it became active. Inasmuch as avoidance responses hinder contact with contingencies, the finding that comparison participants 11 and 12 continued to press the spacebar throughout all the trial-blocks was as expected.

Regarding the first exploratory question about the facilitative effect of verbal prompts, the fact that comparison participants (11 and 12) never ceased to avoid is partly explained by the lack of “challenge questions” acting as verbal *prompts* (cf. Skinner, 1957; Skinner, 1953/1965), which would have aided contacting the “extinct” stimulus-tone contingency. Such a manner of leading to response extinction and its transfer, expands traditional modes of provoking this effect both in basic research and its application. For example, in Dougher et al. (1994) study extinction was achieved by presenting B1 in the absence of shock until the conditioned response had subsided altogether. This is akin to the exposure and flooding techniques typical of behaviour therapy. In the former, the principle is contact with contingencies by an eventual decrease or fluctuation in response rate (as a function of previous reinforcement rate). In the latter, the strategy relies partly on the physiological response depletion (and its subsequent homeostatic tendency) as well as habituation, but also on the presumed “re-conditioning” that occurs once the person is still exposed to the fearful context without experiencing autonomic responses. The use of verbal prompts in the present study resembles the manner in which therapists challenge their patients and encourage them to “find the evidence” of their evaluations about a particular feared situation.

Human operant behaviour is significantly influenced by verbal instructions (see Baron & Galizio, 1983, for a review) and several studies suggest that sensitivity to natural contingencies is often overridden by instructional control. Equally, insensitivity is shown to changes in the programmed consequences of responding and schedule differences. For example as in *fixed-interval* (FI) vs. *fixed-ratio* (FR), or *variable-interval* (VI) vs. *variable-*

ratio (VR) (e.g., Ader & Tatum, 1961; DeLuca & Holborn, 1985; DeLuca & Holborn, 1992; Galizio, 1979; Harzem, Lowe, & Bagshaw, 1978; Hayes, Brownstein, Haas & Greenway, 1986; Joyce & Chase, 1990; Kudadjie-Gyamfi & Rachlin, 2002; Matthews, Shimoff, Catania, & Sagvolden, 1977; Shimoff, Catania, & Matthews, 1981; Weiner, 1970; Wulfert, Greenway, Farkas, Hayes, Dougher, 1994). This has led to rigorous attention to the implementation of instructions, for they can not only affect contingency-shaped behaviour, but also disguise schedule sensitivity (Hayes, Brownstein, Zettle, Rosenfarb, & Korn, 1986).

Based on Ribes-Iñesta's (2000) conceptualization of instructions/rules (but see the inconclusive debate on this issue: O'Hora & Barnes-Holmes, 2001; Ribes-Iñesta, 2001; O'Hora & Barnes-Holmes, 2001a; Ribes-Iñesta, 2001a; see also Ribes, 1999), instructions were given prior to each of the tasks, as well as a standardized method to record the rules they generated via buttons with categorical semantic labels (phase 5 and 6). An unexpected misinterpretation of the instructions, similar to that reported by Martinez and Tamayo (2005), might have been responsible for the performance of participant 15 whose responding may have been guided by the "acquisition of points", thus preventing him from responding in accordance to the "relevant" contingencies of reinforcement. Also, the word "several" was taken to mean "a few" (i.e., "three or four presses") by participants 7, 9 and 12 (all from arrangement B) who required re-set to classical conditioning trial-blocks –reiterating the variability and subtlety of the control that language can exert.

The second exploratory question of this study revolves around the congruence between the participant's verbal estimations and avoidance responses. From the observed interaction, especially from arrangement A, it is tempting to point up a tendency of the estimations to be more sensitive to the "corroborative" contingencies (i.e., change occurred more rapidly) than the avoidance response. However, with such scant data there is a risk of drawing spurious conclusions.

To further determine unintended effects caused by the instructions, or whether the observed behaviour was in fact exclusively controlled by these, participant 15 received experimental arrangement A with the difference that only minimal instructions were provided. In this manner, such a variable was explored and it could be said that the formation of equivalence relations was significantly fostered by the (non-verbal) contingencies of reinforcement. In support of this, participant 15 performed like the other participants with regard to the execution of the tasks and sensitivity to the programmed contingencies. Alternatively to the control exerted by the possibly misleading instructions, the extinguishing effect over the "corroborative response" could have been a function of the –punishing– "negative feedback" obtained during this particular block (even if the contingency also implied the key-pressing response).

Although for participants in arrangement B, only the stimulus being directly paired with the tone elicited key responses, when these participants were questioned at the end, they did report the relation among stimuli and "felt like responding" to them, and some others "sensed the connection". Perhaps the reported relation among stimuli (viz., covert response) could have been confirmed by electrophysiological measurements such as in Augustson and Dougher (1997) and Roche and Barnes (1997), despite lacking overt behavioural manifestation.

It is of growing consensus that anxiety disorders are generally characterized by a tendency to avoid, suppress or escape aversive contexts, including emotional states or cues that may evoke them (e.g., Barlow, 2002; Barlow, Allen, & Choate, 2004; Rosen & Schulkin, 1998). In spite of this, addressing overt avoidance is not only more pragmatic for psychotherapeutic practices (instead of autonomic measurements and referring to inferred tendencies) but also clinically relevant within the context of psychopathology, as it is the overt behavioural manifestation that often leads to life impairment, thus affecting the individual’s daily functioning in the different spheres of adjustment (e.g., social, professional, etc).

It could be argued that, when respondent conditioning is established prior to the formation of equivalence classes, and the US is weak, overt transfer of stimulus functions is less likely. In their first experiment, Dougher et al. (1994) considered that the failure of one participant to show conditioning, and another to show transfer, was due to the relatively weak shock that was used as the US. Nevertheless, despite the use in the present study of an arguably less noxious aversive stimulus, such an argument receives no support from the performance of all the participants from arrangement A, all of whom demonstrated both conditioning effects and their transfer across equivalence class members. Perhaps, such differences are due to the aversiveness of the stimulation used and/or the measurements taken.

The nature of the US employed in the present study merits a phenomenological note nonetheless. After their participation, all participants were asked if they found the tone aversive. Despite the fact that the majority confirmed this assumption about the nature of the stimulation (and the generated emotional reaction), participant 1 and 7 reported indifference (e.g., according to participant 7: “it was a bit annoying but bearable”). In such cases, it would be logical to interpret the transfer as discriminative (i.e., cued by the instructions during avoidance training) rather than motivational stimulus functions (i.e., derived aversive properties).

In extrapolating these findings to clinical settings, the degree to which contingencies undergo extinction, and the degree of transfer across equivalent classes, calls for important considerations. For instance, in equivalence studies dealing with extinction, the target stimulus tends to be the one with direct history of parity with the eliciting –or discriminative–

functions. Would the extinction effect be partial if the extinction occurred with equivalence class members that hold a *derived* relation to the tone? In other words, would extinction be limited to the indirectly related stimuli if one of these were targeted instead? Although an empirical exploration of this hypothesis is still pending, the findings of this paper suggest this possibility (cf. Roche, Kanter, Brown, Dymond, & Fogarty, 2008).

In this sense, in order for a particular treatment which requires response extinction to be effective, the non-adaptive response to be targeted would have to be the one holding the contingency acquired through direct contact –essentially but not exclusively. Namely, in the context of this experiment the contingency to be "treated" through a process of extinction would have to prioritise exposure to B1 stimulus.

Concerning "relapse", the phenomenon of *spontaneous recovery* (e.g., Skinner, 1938/1991) referring to eliciting properties, could speculatively be explained by the fact that as long as B1's stimulus functions remain to some extent, more associations are forthcoming (cf. Wilson & Hayes, 1996). If responses to B1 endure, even though responses to other class members have undergone extinction, there is a possibility that B1 could enter in other networks and potentially form separate equivalence classes; thus contributing to the maintenance of the response of relevance or simply laying the conditions for other "equivalences" (viz., response generalization), unless the contingency operating on B1 stimulus is systematically tackled.

Conclusion

In line with other studies (e.g., Augustson & Dougher, 1997), the data reported here corroborate the transfer of avoidance functions via equivalence relations. In addition, this study has also demonstrated: a) the decline of response strength of derived stimulus functions across equivalence classes during extinction; b) derived avoidance responses undergo extinction somewhat more readily than directly conditioned avoidance responses; c) verbal prompts facilitate contact with contingencies, thus contributing to sensitivity insofar as these are contingency-specific; and d) verbal estimations about the contingencies experienced (viz., rules) do not accurately correspond to the non-verbal avoidance responses, an aspect which demands more thorough experimental exploration.

Finally, there will inevitably be pitfalls in extrapolating basic research findings to applied settings. Under laboratory conditions, one controls for pre-existing equivalence classes by using unfamiliar stimuli and training the intended relations among them through arranged contingencies of differential reinforcement. Extra-laboratory situations entail a vast array of “historical” and ongoing equivalence classes in matured individuals, with language serving as an “equivalencing” vehicle. For example, if replications of the present study were made with clinical populations, would “anxious people” require more extinction trials despite the constant negative feedback? (hence showing more insensitivity to contingencies?), would they need more “invalidating consequences” before a change in the contingency occurred? Research employing clinical populations in this area is scarce, and more data could contribute to understanding alternative ways in which transferred functions may be sensitive to being broken by means other than classical extinction and operant conditioning (see Dymond & Roche, 2009; Hayes, Strosalhl, & Wilson, 1999; Hayes & Strosahl, 2004; Woods & Kanter, 2007; for empirically oriented clinical applications in accord with developments in this area).

References

Ader, R., & Tatum, R. (1961). Free-operant avoidance conditioning in human subjects. *Journal of the Experimental Analysis of Behavior*, *4*, 275-276.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record*, *60*, 437–462.

Augustson, E. A., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behaviour Therapy and Experimental Psychiatry*, *28*, 181-191.

Augustson, E. A., Dougher, M. J., & Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record*, *50*, 745-770.

Barlow, D. H (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic*. NY: Guilford.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, *35*, 205-230.

Barnes, D., & Roche, B. (1997). Relational frame theory and the experimental analysis of human sexuality. *Applied and Preventive Psychology*, *6*, 117-136.

Barnes, D., Holmes, Y., Smeets, P. M., & Luciano, C. (2004). A derived transfer of mood functions through equivalence relations. *The Psychological Record*, *54*, 95-113.

Baron, A., & Galizio, M. (1983). Instructional control of human operant behavior. *Psychological Record*, *33*, 495-520.

Carrigan, P. F. Jr., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, *58*, 183-204.

DeGrandpre, R. J., & Bickel, W. K. (1993) Stimulus control and drug dependence. *The Psychological Record*, *43*, 651-666.

DeLuca, R. V., & Holborn, S. W. (1985). Effects of a fixed-interval schedule of token reinforcement on exercise with obese and non-obese boys. *Psychological Record*, *35*, 525-533.

DeLuca, R. V., & Holborn, S. W. (1992). Effects of variable-ratio reinforcement schedule with changing criteria on exercise in obese and non-obese boys. *Journal of Applied Behaviour Analysis*, *25*, 671-679.

Dougher, M. J., Augustson, E. M., Markham, M. R., Greenway, D., & 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.

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

Dymond, S., & Roche, B. (2009). A contemporary behavior analysis of anxiety and avoidance. *The Behavior Analyst, 32*, 7-27.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1993). Are stimuli in equivalence classes equally related to each other? *The Psychological Record*, *43*, 85-106.

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior*, *42*, 143-157.

Fields, L., & P. Moss (2008). Formation of partially and fully elaborated generalized equivalence classes. *Journal of the Experimental Analysis of Behavior, 90*, 135–168.

Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior, 31*, 53-70.

Glat, R., Gould, K. Stoddard, L. T., & Sidman, M. (1994). A note on transfer of stimulus control in the delayed-cue procedure: Facilitation by an overt differential response. *Journal of Applied Behavior Analysis*, *27*, 699-704.

Goldiamond, I. (1962). Perception. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology* (pp. 280-340).

Harzem, P., Lowe, C. F., & Bagshaw, M. (1978). Verbal control in human operant behavior. *Psychological Record*, *28*, 405-423.

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior*. (pp. 19-41). Reno, NV: Context Press.

Hayes, S. C., & Strosalhl, K. D. (2004). *A practical guide to acceptance and commitment therapy*. NY: Springer Science+Business Media, Inc.

Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, Multiple Schedules, and Extinction: Distinguishing Rule-Governed from Scheduled-Controled Behavior. *Journal of the Experimental Analysis of Behavior*, *46*, 137-147.

Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior*, *45*, 237-256.

Hayes, S. C., Strosalhl, K. D., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy. An experiential Approach to Behavior Change*. NY: The Guilford Press.

Joyce, J. H., & Chase, P. N. (1990). Effects of response variability on the sensitivity of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, *54*, 251-262

Kudadjie-Gyamfi, E., & Rachlin, H. (2002). Rule-governed versus contingency-governed behavior in a self-control task: effects of changes in contingencies. *Behavioural Processes*, *57*, 29–35.

Martinez, H., & Tamayo, R. (2005). Interactions of contingencies, instructional accuracy, and instructional history in conditional discrimination. *The Psychological Record*, *55*, 633-646.

Matthews, B. A., Shimoff, E., Catania, A. C., & Sagvolden, T (1977). Uninstructed human responding: Sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior*, *27*, 453-467.

O'Hora, D., & Barnes-Holmes, D. (2001). The referencial nature of rules and instructions: A response to instructions, rules, and abstraction: A misconstrued relation by Emilio Ribes-Iñesta. *Behavior and Philosophy*, *29*, 21-25.

O'Hora, D., & Barnes-Holmes, D. (2001a). Stepping up to the challenge of complex human behavior: A response to Emilio Ribes-Iñesta's response. *Behavior and Philosophy*, *29*, 59-60.

Ribes, E. (1999). *Teoría del condicionamiento y lenguaje: Un análisis histórico y conceptual*. Madrid-México: Taurus.

Ribes-Iñesta, E. (2000). Instructions, rules and abstraction: A misconstrued relation. *Behavior and Philosophy*, *28*, 41-55.

Ribes-Iñesta, E. (2001). About persistent conceptual confusion: A response to O'Hora and Barnes-Holmes. *Behavior and Philosophy*, *29*, 27-29.

Ribes-Iñesta, E. (2001a). Stepping down to the foundations is needed to remedy conceptual confusion: A final reply to O'Hora and Barnes-Holmes. *Behavior and Philosophy, 29*, 61-62.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, *67*, 275-301.

Roche, B., Kanter, J. W., Brown, K. R., Dymond, S., & Fogarty, C. C. (2008). A comparison of "direct" versus "derived" extinction of avoidance. *The Psychological Record, 58,* 443-464.

Rosen, J. B., & Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review*, *105*, 325-250.

Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Responsivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior*, *36*, 207-220.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5-13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.). *Analysis and integration of behavioral units* (pp. 213-245). Hilldale, NJ: Erlbaum.

Sidman, M. (1987). Two choices are not enough. *Behaviour Analysis*, *22*, 11-18.

Sidman, M (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, NV: Context Press.

Sidman, M (1992a). Adventitious control by the location of comparison stimuli in conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *58*, 173-182.

Sidman, M. (1994). *Equivalence Relations and Behavior: A research history*. Boston, MA: Authors Cooperative, Inc., Publishers.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*, 127-146

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *The Analysis of Verbal Behavior*, *25*, 5-17.

Sidman, M., & Tailby, W. O. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*, 5-22.

Sidman, M., Rauzing, R., Lazar, R., Cunningham, S., Tailby, W. O., & Carrigan, P. (1982). A research for symmetry in the conditional discriminations of rhesus monkeys, baboons and children. *Journal of the Experimental Analysis of Behavior*, *37*, 23-44.

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.

Sidman, M., Wynne, C. K., Maguire, R. W. & Barnes, T (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior*, *52*, 261-274.

Skinner, B. F. (1957). *Verbal behavior*. NY: Appleton Century Crofts.

Skinner, B. F. (1965). *Science and human behavior*. NY: The free press. (Original work published 1953 by The Macmillan Company).

Skinner, B. F. (1991). *The behaviour of organisms*. Acton, Massachusetts: Copley Publishing Group. (Original work published 1938).

Steele, D., & Hayes, C. S. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, *56*, 519-555.

Weiner, H. (1970). Human behavioral persistence. *Psychological Record*, *20*, 445-456.

Wilson K. G., & Hayes S. C. (1996). Resurgence of derived stimulus relations. *Journal of the Experimental Analysis of Behavior*, *66*, 267-281.

Woods, D. W., & Kanter, J. W. (2007). *Understanding behavior disorders: A contemporary behavioral perspective*. Reno, NV: Context Press.

Wulfert, E., Greenway, D. E., Farkas, P., Hayes, S. C., & Dougher, M. J. (1994). Correlation between self-reported rigidity and rule-governed insensitivity to operant contingencies. *Journal of Applied Behavior Analysis*, *27*, 659-671.

APPENDIX

Appendix A.

Appendix A. Order of stimuli presentation during the different experimental trials for each participant. The "shadowed" characters correspond to class 3 stimuli that appeared interspersed with the others.

Particip.	**Trials**	**Order of stimuli presentation - *Arrangement A***
P1	Trans:	D3, B3, C2, C3, B1, D2, D1, B2, C1, B2, D2, B1, D1, C1, D3, C2, B3, C3
	Estim:	B2, D3, C1, D1, B3, C2, D2, B1, C3
	Corrob:	D2, D1, C1, B1, B3, D3, C3, B2, C2, C1, B1, D2, B2, C2, D1, C2, B2, D2, D1, D3, B1, C1
P2	Trans:	C3, B3, D3, B1, C1, D1, B2, C2, D2, D1, C1, D3, B2, B1, B3, C2, D2, C3
	Estim:	C1, D3, B2, C3, D2, C2, B3, B1, D1
	Corrob:	D3, C1, C3, B3, D1, B2, D2, C2, B1, C1, D2, C3, B3, D1, C2, D3, B2, B1, D2, B1, B3, B2, C2, D1, C3, C1, D3
P3	Trans:	C1, B2, C2, C3, D2, D1, B1, B3, D3, D3, B3, B2, D2, B1, D1, C1, C3, C2
	Estim:	C3, B1, D2, D1, C1, C2, B3, D3, B2
	Corrob:	D2, D3, C1, C3, C2, D1, B3, B2, B1, D2, B2, C3, B1, D1, C2, D3, B3, C1, C3, C2, D2, D1, B1, D3, C1, B3, B2
P11*c*	Trans:	B1, B2, B3, D1, C1, C2, C3, D2, D3, B1, C2, B2, D1, C1, B3, D2, C3, D3
	Estim:	B2, D1, B3, D2, B1, D3, C3, C1, C2
	Extinc:	B3, B1, D1, D3, C3, C2, D2, B2, C1, B1, B2, C2, B3, D2, D1, C1, D3, C3, B1, D2, B2, C2, C3, D1, B3, D3, C1
		Arrangement B
P7	Trans:	C1, B3, B2, C2, C3, D2, D3, B1, D1, D3, B1, B3, C3, D1, D2, B2, C1, C2
	Estim:	B2, D1, D2, C1, C3, B1, C2, D3, B3
	Corrob:	D1, C2, D3, B2, B3, D2, C1, C3, B1, D1, C2, B2, D2, B1, C1, B3, C3, D3, D2, B2, C1, C2, D1, B1, C3, D3, B3
P8	Trans:	B3, D2, C3, B1, C1, B2, C2, D1, D3, D3, D1, B3, C1, D2, C2, B1, C3, B2
	Estim:	D1, C2, C1, D3, B1, C3, D2, B2, B3
	Corrob:	C3, B2, B1, D3, C1, D2, D1, C2, B3, D2, C3, D1, D3, B2, C2, B1, B3, C1, D3, C1, B2, D2, B1, C3, B3, D1, C2
P9	Trans:	B2, D1, B3, B1, D3, C1, C3, C2, D2, C1, B1, C2, D2, C3, B3, D1, B2, D3
	Estim:	B1, B2, D1, B3, D3, C1, D2, C3, C2
	Corrob:	C2, B2, D3, C3, B3, C1, B1, D2, D1, C3, B1, B3, B2, C1, C2, D1, D3, D2, B3, C2, D2, C3, D1, D3, B1, B2, C1
P12*c*	Trans:	B1, D2, C1, D1, B2, C2, C3, B3, D3, B1, B3, C2, B2, C1, D1, D3, D2, C3
	Estim:	B3, D2, C1, D1, C3, C2, B2, B1, D3
	Extinc:	B1, C1, C3, B2, D3, B3, C2, D1, D2, B3, C1, D3, D1, B2, C2, D2, B1, C3, D1, B3, C3, B2, C2, D3, D2, C1, B1
		Participant 15 (min instructions)
P15	Trans:	B1, C3, D1, C2, C1, D2, B2, B3, D3, B1, D2, D3, B2, B3, C1, C2, C3, D1
	Estim:	C1, C3, B3, D1, B1, B2, D3, D2, C2
	Corrob:	C1, B2, C2, D3, B1, C3, B3, D2, D1, D1, C2, C3, C1, D2, D3, B2, B1, B3, D2, D3, C3, D1, C2, B2, C1, B1, B3

Author Note

Part of this research fulfilled a requirement for the first author's MSc in Psychology at the University of East London. The computer software was developed by Tony Leadbetter, senior technical officer at the University. Correspondence and reprint requests should be addressed to Santiago Garcia Guerrero: santiago.garciadefilippis@gmail.com

Tables

Table 1. Permutations of relations among stimuli and their trial-type presentation.

	TASK/ BLOCK	SAMPLE	CORRECT	INCORRECT				TASK/ BLOCK	SAMPLE	CORRECT	INCORRECT		
TRAINED / BASELINE		A1	B1	B2	B3	B4	*TRANSITIVITY*		B1	C1	C2	C3	C4
	AB	A2	B2	B1	B3	B4		**BC**	B2	C2	C1	C3	C4
		A3	B3	B1	B2	B4			B3	C3	C1	C2	C4
		A1	C1	C2	C3	C4			C1	B1	B2	B3	B4
	AC	A2	C2	C1	C3	C4		**CB**	C2	B2	B1	B3	B4
		A3	C3	C1	C2	C4			C3	B3	B1	B2	B4
		A1	D1	D2	D3	D4			C1	D1	D2	D3	D4
	AD	A2	D2	D1	D3	D4		**CD**	C2	D2	D1	D3	D4
		A3	D3	D1	D2	D4			C3	D3	D1	D2	D4
SYMMETRY		B1	A1	A2	A3	A4			D1	C1	C2	C3	C4
	BA	B2	A2	A1	A3	A4		**DC**	D2	C2	C1	C3	C4
		B3	A3	A1	A2	A4			D3	C3	C1	C2	C4
		C1	A1	A2	A3	A4			B1	D1	D2	D3	D4
	CA	C2	A2	A1	A3	A4		**BD**	B2	D2	D1	D3	D4
		C3	A3	A1	A2	A4			B3	D3	D1	D2	D4
		D1	A1	A2	A3	A4			D1	B1	B2	B3	B4
	DA	D2	A2	A1	A3	A4		**DB**	D2	B2	B1	B3	B4
		D3	A3	A1	A2	A4			D3	B3	B1	B2	B4

REFLEXIVITY

TASK/ BLOCK	Sample	Correct	Incorrect			TASK/ BLOCK	Sample	Correct	Incorrect		
	A1	A1	N5	N6	N7		B1	B1	N5	N6	N7
AA	A2	A2		*N Stimuli*		**BB**	B2	B2		*N Stimuli*	
	A3	A3		*Randomized*			B3	B3		*Randomized*	
	C1	C1	N5	N6	N7		D1	D1	N5	N6	N7
CC	C2	C2		*N Stimuli*		**DD**	D2	D2		*N Stimuli*	
	C3	C3		*Randomized*			D3	D3		*Randomized*	

Table 2
Number of incorrect selections each participant had over the total amount of trials per block type (First four participants).

Participant	***Baseline***	**AB**	**AC**	**AD**						
1	5/194	2/77	1/61	2/56						
2	3/374	3/191	0/93	0/90						
3	11/236	5/97	1/61	5/78						
7	23/261	13/96	4/82	6/83						
	Mixed Baseline	**AB**	**AC**	**AD**						
1	1/27	0/9	1/9	0/9						
2	0/54	0/18	0/18	0/18						
3	0/27	0/9	1/9	0/9						
7	0/54	0/18	0/18	0/18						
	Symmetry	**BA**	**CA**	**DA**						
1	0/36	0/12	0/12	0/12						
2	0/36	0/12	0/12	0/12						
3	0/36	0/12	0/12	0/12						
7	0/102	0/34	0/34	0/34						
	Transitivity	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**			
1	0/72	0/12	0/12	0/12	0/12	0/12	0/12			
2	0/72	0/12	0/12	0/12	0/12	0/12	0/12			
3	0/72	0/12	0/12	0/12	0/12	0/12	0/12			
7	37/216	7/36	10/36	1/36	1/36	10/36	8/36			
	Mixed S&T	**BA**	**CA**	**DA**	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**
1	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
2	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
3	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
7	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
	Reflexivity	**AA**	**BB**	**CC**	**DD**					
1	0/12	0/3	0/3	0/3	0/3					
2	1/12	0/3	0/3	1/3	0/3					
3	0/12	0/3	0/3	0/3	0/3					
7	9/12	2/3	3/3	2/3	2/3					
	Post-experiment Mixed S&T	**BA**	**CA**	**DA**	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**
1	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
2	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
3	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
7	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3

Table 2 (Continuation)
Number of incorrect selections over the total of trials per block type (Five other participants).

Participant	***Baseline***	**AB**	**AC**	**AD**					
8	20/270	14/109	4/86	2/75					
9	28/236	14/91	12/89	2/56					
11	13/112	8/45	3/35	2/32					
12	33/312	19/134	8/92	6/86					
15	37/229	31/155	4/41	2/33					
	Mixed Baseline	**AB**	**AC**	**AD**					
8	0/27	0/9	0/9	0/9					
9	1/27	0/9	0/9	1/9					
11	0/27	0/9	0/9	0/9					
12	0/54	0/18	0/18	0/18					
15	0/27	0/9	0/9	0/9					
	Symmetry	**BA**	**CA**	**DA**					
8	2/36	1/12	1/12	0/12					
9	2/36	0/12	1/12	1/12					
11	0/36	0/12	0/12	0/12					
12	8/108	8/36	0/36	0/36					
15	1/36	1/12	0/12	0/12					
	Transitivity	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**		
8	2/72	1/12	0/12	0/12	0/12	1/12	0/12		
9	3/144	0/24	2/24	0/24	0/24	0/24	1/24		
11	2/72	0/12	1/12	0/12	1/12	0/12	0/12		
12	2/72	1/12	0/12	0/12	0/12	0/12	1/12		
15	2/72	0/12	0/12	0/12	1/12	0/12	1/12		
	Mixed S&T	**BA**	**CA**	**DA**	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**
8	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
9	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
11	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
12	1/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3
15	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
	Reflexivity	**AA**	**BB**	**CC**	**DD**					
8	0/12	0/3	0/3	0/3	0/3					
9	0/12	0/3	0/3	0/3	0/3					
11	0/12	0/3	0/3	0/3	0/3					
12	1/12	0/3	0/3	0/3	1/3					
15	0/12	0/3	0/3	0/3	0/3					
	Post-experiment Mixed S&T	**BA**	**CA**	**DA**	**BC**	**CB**	**CD**	**DC**	**BD**	**DB**
8	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
9	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
11	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
12	1/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3
15	0/27	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3

Table 3. Shows arrangement A participants' performance during the 1st and 2nd trial *tests for transfer*, 3rd trial for *estimation* and 4th, 5th and 6th trial for *corroboration*. Comparison participant 11 did not receive corroboration trials.

Participant 1

	Test 1	**Test** 2	**Est** 3	**Corroboration** 4	**Corroboration** 5	**Corroboration** 6
B1	42	31	28 (1)	27 (Y) –	69 (Y) –	0 (N) +
C1	27	30	35 (1)	14 (N) +	0 (N) +	0 (N) +
D1	37	33	25 (1)	28 (Y) –	0 (N) +	0 (N) +
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 2

	Test 1	**Test** 2	**Est** 3	**Corroboration** 4	**Corroboration** 5	**Corroboration** 6
B1	26	22	22 (1)	17 (Y) –	23 (Y) –	0 (N) +
C1	12	21	15 (2)	24 (Y) –	0 (N) +	0 (N) +
D1	24	18	18 (2)	13 (Y) –	0 (N) +	0 (N) +
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 3

	Test 1	**Test** 2	**Est** 3	**Corroboration** 4	**Corroboration** 5	**Corroboration** 6
B1	20	19	28 (2)	26 (Y) –	0 (N) +	0 (N) +
C1	7	18	22 (2)	19 (Y) –	0 (N) +	0 (N) +
D1	17	19	29 (2)	29 (Y) –	0 (N) +	0 (N) +
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 11 (comparison)

	Test 1	**Test** 2	**Est** 3	**Est in Extinction** 4	**Est in Extinction** 5	**Est in Extinction** 6
B1	9	11	8 (1)	7 (1)	9 (1)	9 (2)
C1	11	11	6 (2)	9 (2)	7 (2)	8 (2)
D1	10	12	6 (2)	8 (2)	0 (3)	9 (2)
B2	0	0	0 (3)	0 (3)	0 (3)	0 (3)
C2	0	0	0 (4)	0 (4)	0 (4)	0 (3)
D2	0	0	0 (4)	0 (4)	0 (3)	0 (3)

The stand-alone numbers correspond to the spacebar presses for each stimulus; the numbers within brackets in trial 3 correspond to the estimation buttons: 1 = "definitely happening", 2 = "probably happening", 3 = "definitely not happening", 4 = "probably not happening"; the letters within brackets in trials 4, 5 and 6 correspond to the estimation options: (Y) = "Happened", (N) = "Not happened"; and the plus + or minus – indicate whether the feedback received when corroborating their estimations was positive or negative. The dot • means that the participant did not corroborate.

Table 4. Shows arrangement B participants' performance during the 1st and 2nd trial *tests for transfer*, 3rd trial for *estimation* and 4th, 5th and 6th trial for *corroboration*. Comparison participant 12 did not received corroboration trials. (data conventions as in the previous table)

	Test		Est	Corroboration		
	1	2	3	4	5	6
B1	22	16	13 (1)	15 (Y) –	16 (Y) –	0 (N) ·
C1	0	17	13 (2)	9 (N) +	0 (N) ·	0 (N) ·
D1	19	19	16 (2)	16 (Y) –	0 (N) ·	0 (N) ·
B2	0	0	0 (3)	0 (N) ·	0 (N) ·	0 (N) ·
C2	0	0	0 (3)	0 (N) +	0 (N) ·	0 (N) ·
D2	0	0	0 (3)	0 (N) ·	0 (N) ·	0 (N) ·

Participant 7

	Test		Est	Corroboration		
	1	2	3	4	5	6
B1	0	20	1 (1)	47 (Y) –	44 (N) +	0 (N) +
C1	0	0	0 (4)	0 (N) +	0 (N) +	0 (N) +
D1	0	0	0 (2)	0 (N) +	0 (N) +	0 (N) +
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 8

	Test		Est	Corroboration		
	1	2	3	4	5	6
B1	9	11	11 (2)	0 (Y) –	0 (Y) –	0 (N) +
C1	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D1	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 9

	Test		Est	Est in Extinction		
	1	2	3	4	5	6
B1	12	16	18 (1)	15 (1)	15 (1)	14 (1)
C1	0	0	0 (4)	0 (2)	0 (2)	0 (4)
D1	0	0	0 (4)	0 (4)	0 (4)	0 (4)
B2	0	0	0 (3)	0 (3)	0 (3)	0 (3)
C2	0	0	0 (3)	0 (3)	0 (3)	0 (3)
D2	0	0	0 (3)	0 (3)	0 (3)	0 (3)

Participant 12 (comparison)

Table 5. Shows participant 15 (minimal instructions) data during the tests for transfer, estimation and corroboration trial-blocks.

	Test		Est	Corroboration		
	1	2	3	4	5	6
B1	19	22	21 (1)	20 (Y) –	27 (Y) –	24 (Y) ·
C1	23	20	30 (1)	21 (Y) –	28 (Y) ·	35 (Y) ·
D1	21	20	50 (1)	26 (Y) –	25 (Y) ·	28 (Y) ·
B2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
C2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +
D2	0	0	0 (3)	0 (N) +	0 (N) +	0 (N) +

Participant 15

Figure Captions

Figure 1. Arbitrary stimuli classes and their members.

Figure 2. Order of the experimental phases. Phase 1 took place after phase 3 for participants assigned to arrangement B. Two comparison participants received an extended phase 5 instead of phase 6. Another completed the entire experiment (arrangement A) with minimal instructions.

Figure 3. Diagram showing the sequence flow of the experimental phase fostering equivalence relations. Reflexivity tests are not included. Every “Yes” in the diagram resulted in a text box superimposed on the screen saying “Block trial completed”.

Figures

Figure 1.

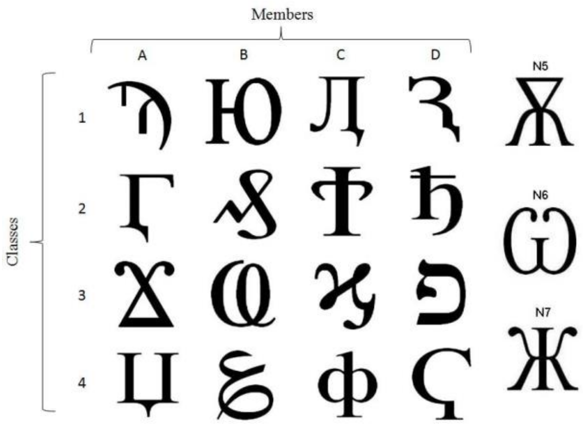


Figure 2.

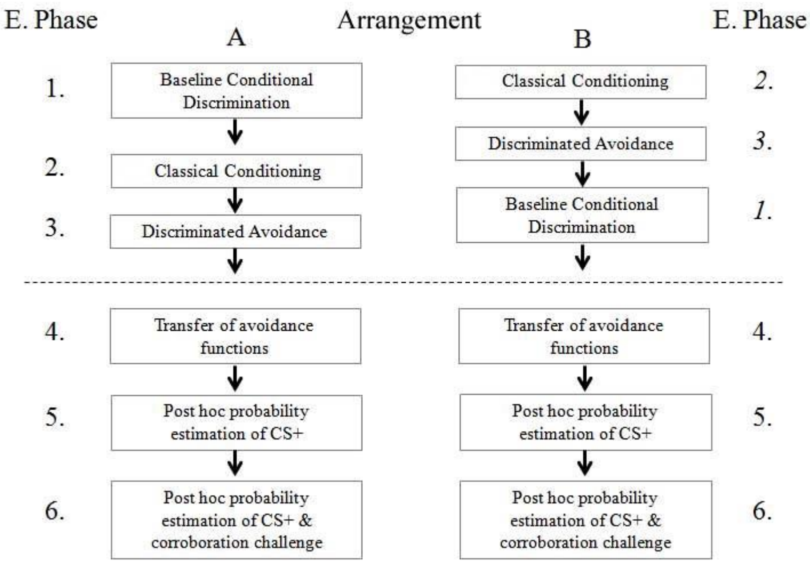
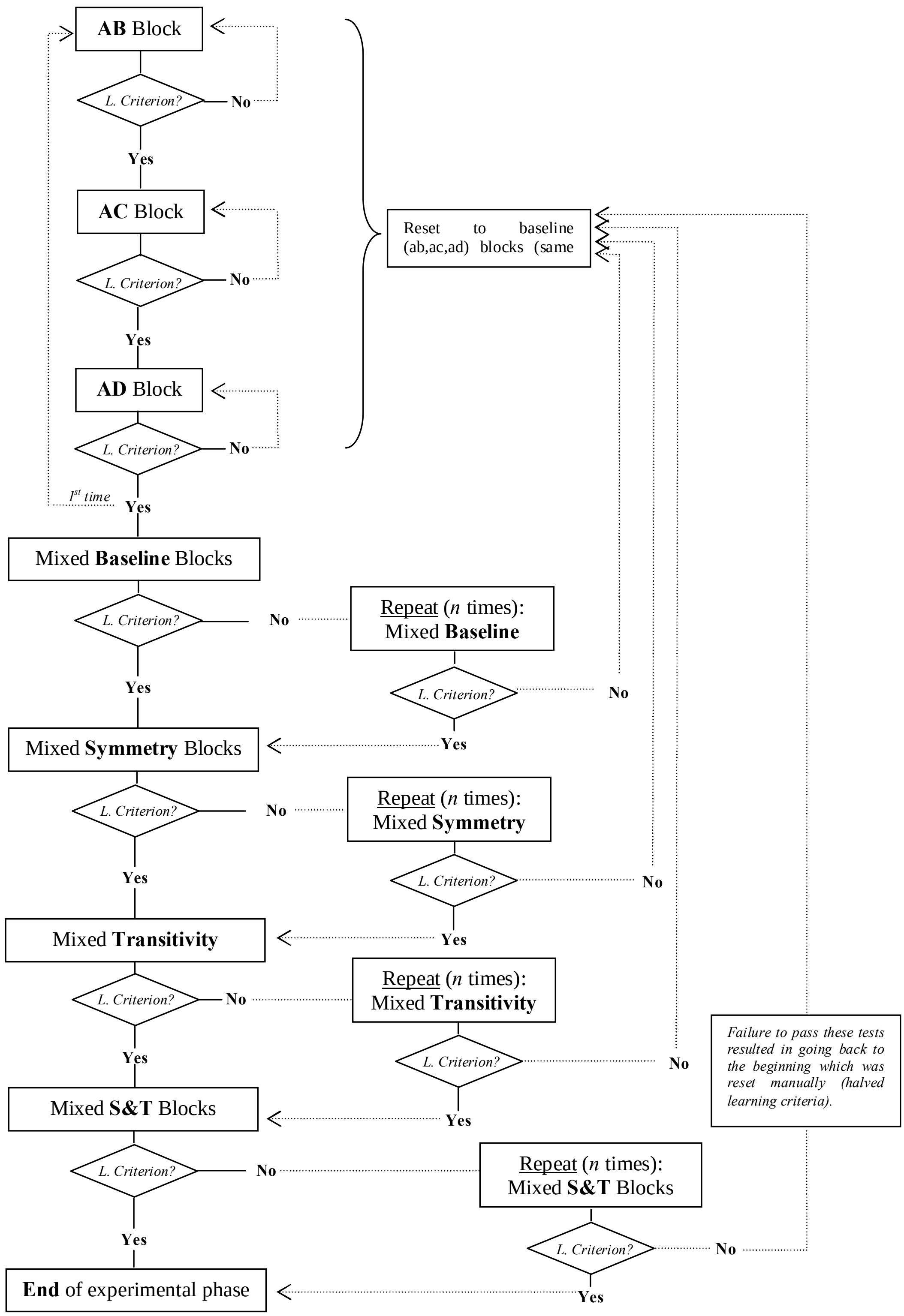

E. Phase
Arrangement
E. Phase
A
B
1.
Baseline Conditional Discrimination
2.
Classical Conditioning
3.
Discriminated Avoidance
Classical Conditioning
2.
Discriminated Avoidance
3.
Baseline Conditional Discrimination
1.
4.
Transfer of avoidance functions
Transfer of avoidance functions
4.
5.
Post hoc probability estimation of CS+
Post hoc probability estimation of CS+
5.
6.
Post hoc probability estimation of CS+ & corroboration challenge
Post hoc probability estimation of CS+ & corroboration challenge
6.

Figure 3.



AB Block
L. Criterion?
No
Yes
AC Block
L. Criterion?
No
Yes
AD Block
L. Criterion?
No
1st time
Yes
Reset to baseline (ab,ac,ad) blocks (same
Mixed Baseline Blocks
L. Criterion?
No
Yes
Repeat (n times): Mixed Baseline
L. Criterion?
No
Yes
Mixed Symmetry Blocks
L. Criterion?
No
Yes
Repeat (n times): Mixed Symmetry
L. Criterion?
No
Yes
Mixed Transitivity
L. Criterion?
No
Yes
Repeat (n times): Mixed Transitivity
L. Criterion?
No
Yes
Mixed S&T Blocks
L. Criterion?
No
Yes
Repeat (n times): Mixed S&T Blocks
L. Criterion?
No
Yes
End of experimental phase
Failure to pass these tests resulted in going back to the beginning which was reset manually (halved learning criteria).