Accepted Manuscript

Authors: Simon Dymond, Michael W. Schlund, Bryan Roche, Robert Whelan, Jennifer Richards, Cara Davies

Running head: INFERRED THREAT AND SAFETY

Inferred Threat and Safety: Symbolic Generalization of Human Avoidance Learning

Simon Dymond[1], Michael W. Schlund[2], Bryan Roche[3], Robert Whelan[4], Jennifer Richards[1], and

Cara Davies[1]

[1] Department of Psychology, Swansea University, Singleton Park, Swansea SA2 8PP, United

Kingdom (s.o.dymond@swansea.ac.uk).

[2] Departments of Behavioural Psychology, Kennedy Krieger Institute, 707 North Broadway,

Baltimore, MD 21205, USA and Behaviour Analysis, University of North Texas, PO Box

310919, Denton, TX 76203, USA (schlund@kennedykrieger.org).

[3] Department of Psychology, National University of Ireland, Maynooth, Co. Kildare, Ireland

(Bryan.T.Roche@nuim.ie).

[4] Trinity Centre for Bioengineering, Printing House, Trinity College Dublin, Dublin 2, Ireland

(robert.whelan@tcd.ie).

Corresponding author: Dr Simon Dymond, Department of Psychology, Swansea

University, Singleton Park, SA2 8PP, United Kingdom. Tel: +44 1792 295602, Fax: +44 1792

295679 (s.o.dymond@swansea.ac.uk).

Abstract

Symbolic generalization of avoidance may underlie the etiology and maintenance of anxiety disorders. The aim of the present study was to demonstrate inferred threat-avoidance and safety (non-avoidance) behaviours that occur in the presence of stimuli indirectly related to learned threat and safety cues. A laboratory experiment was conducted involving two symbolic stimulus equivalence relations consisting of three physically dissimilar stimuli (avoidance cues: AV1-AV2-AV3 and neutral cues: N1-N2-N3). During avoidance learning involving aversive images and sounds, a key-press avoidance response was trained for one member of one of the relations (AV2) and non-avoidance for another (N2). Inferred threat and safety behaviour and ratings of the likelihood of aversive events were tested with presentations of all remaining stimuli. Findings showed a significantly high percentage of avoidance to both the learned and inferred threat cues and less avoidance to both the learned and inferred safety cues. Ratings in the absence of avoidance were high during training and testing to threat cues and low to safety cues and lower in the presence of avoidance. Implications for associative and behavioural accounts of avoidance, and modern therapies for anxiety disorders are discussed.

Keywords: threat; safety; avoidance; symbolic generalization; stimulus equivalence.

Safety behaviours occur in specific stimulus situations and involve overt or covert avoidance of potentially fearful or threatening outcomes (Rachman, Radomsky, & Shafran, 2008; Salkovskis, 1991; Salkovskis, Clark, Hackman, Wells, & Gelder, 1999). The recurrence of safety behaviours is considered to lead to the maintenance of anxiety through a failure to learn accurate threat-relevant information in different situations and from an overall diminution of approach behaviours. For instance, an individual diagnosed with social phobia who avoids all social situations not only learns an increased attentional salience for avoidance-evoking, safety-relevant cues but also experiences fewer opportunities to disconfirm negative beliefs (Salkovskis, 1991). According to cognitive accounts, safety behaviours may often be mistakenly attributed as preventing the occurrence of the feared outcome and lead to an updating of erroneous threat beliefs (Salkovskis, 1991). It is usually the target of cognitive behaviour therapy to undertake exposure in a context in which safety behaviours are precluded, leading to a reduction in both fear and threat expectancies.

Avoidance has a crucial role in safety behaviours and serves to maintain fear and anxiety. Laboratory research has shown that avoidance is learned when a response terminates a warning signal such as a light or tone that precedes an unconditioned stimulus (US) such as electric shock and thus prevents occurrence of the US. By virtue of Pavlovian conditioning, the warning signal comes to function as a conditioned stimulus (CS) and elicit fear or anxiety because of its prior relationship with the US when the avoidance response is not made. Operant conditioning occurs when the avoidance response is made in the presence of the CS and leads to omission of the US.

Lovibond (2006) developed an integrated cognitive expectancy model of avoidance learning. The main feature of the model is that participants are said to develop expectancies or propositional knowledge that a particular CS is followed by a particular US and that performing the avoidance response results in the omission of the expected US. Crucially, according to this

account, avoidance learning is said to result from controlled, reasoning processes in which the

expected outcomes of responding and not responding are compared. In this way, "avoidance

interacts with anxiety through the mediating process of expectancy of the aversive outcome"

(Lovibond, 2006, p.126). Several laboratory studies have provided support for the key features of

the expectancy model of avoidance learning (Declercq & De Houwer, 2008, 2009a, 2011;

Lovibond, Saunders, Weidemann, & Mitchell, 2008; Lovibond, Mitchell, Minard, Brady, &

Menzies, 2009; Ly & Roelofs, 2009). Despite these advances, an important challenge remains in

how to conceptualize avoidance that occurs not only in the presence of the CS but also in the

presence of stimuli indirectly related to the CS or US.

A potentially promising avenue of investigation may be to employ verbal relations, such

as stimulus equivalence, in the experimental analysis of clinical behaviour (Dymond & Roche,

2009). Research on verbal relations has shown that when language-able humans are taught a

series of interrelated discriminations involving physically dissimilar (arbitrary) stimuli, the

stimuli involved often become related to each other in ways not explicitly trained (Sidman,

1994). To illustrate, if choosing Stimulus X in the presence of Stimulus A is taught (i.e., A-X),

and choosing Stimulus Y in the presence of Stimulus A (i.e., A-Y) is also taught, it is likely that

untrained relations will emerge between X and A, Y and A ("symmetry"), X and Y, and Y and X

("combined symmetry and transitivity", or "equivalence"), in the absence of any feedback. When

these relations have emerged, a stimulus equivalence relation is said to have formed among the

relata (Hayes & Hayes, 1992; Sidman, 1994). These untrained verbal relations may help to

explain the indirect or *inferred* control of avoidance behaviour that is often seen in the anxiety

disorders.

Supporting evidence for this approach comes from a study that demonstrated inferred

threat-avoidance and safety learning (Augustson & Dougher, 1997). Participants were first

trained and tested for the formation of stimulus equivalence relations consisting entirely of abstract, arbitrary stimuli (AV1-AV2-AV3-AV4 and N1-N2-N3-N4; note that AV refers to cues from the class of avoidance stimuli, and N refers to neutral cues). Next, in a differential conditioning procedure, one stimulus (AV2) was followed by shock and another (N2) was not. During the subsequent avoidance-learning phase, AV2 was followed by shock unless a fixed-ratio 20-response requirement was met, in which case AV2 was removed from the screen and the scheduled shock omitted. The transfer of this threat-avoidance responding was then tested with presentations of stimuli that had not been present during the avoidance-learning phase. Findings showed that all participants emitted the threat-avoidance response to AV3 and AV4 (indirectly related to AV2) and not to N3 and N4 (indirectly related to N2). Further studies have replicated and extended this basic effect (Dymond, Roche, Forsyth, Whelan, & Rhoden, 2007, 2008; Gannon, Roche, Kanter, Forsyth, & Linehan, in press; Roche, Kanter, Brown, Dymond, & Fogarty, 2008).

In a recent study, Declercq and De Houwer (2009b) used a sensory pre-conditioning procedure to investigate the indirect control of avoidance. Declercq and De Houwer (2009b) first paired two neutral stimuli A and B with two other neutral stimuli K and L. Next, A and B were each followed by a unique US (a red 'X' and a red 'Y', respectively, resulting in money loss). In the avoidance learning phase, participants learned to make one avoidance response (R1) in the presence of A and another avoidance response (R2) in the presence of B that both resulted in omission of the US. The critical test phases involved presenting K and L, and several control stimuli, and asking participants to perform one of the two responses and to rate the likelihood of the US appearing. Findings showed that participants both selected avoidance responses and formed expectancies based on the integration of knowledge concerning the US following K and L

and the avoidance of the US by making the appropriate avoidance response, thus supporting

Lovibond's (2006) account.

The aim of the present study was to extend the findings of both Augustson and Dougher

(1997) and Declercq and De Houwer (2009b) by testing whether the predictions made by

Lovibond's (2006) expectancy model of avoidance could explain indirect acquisition of

avoidance. The study will address three main issues. First, Lovibond's model would presumably

explain the emergence of avoidance behaviour in the presence of AV3 and AV4 as resulting from

the expectancy that a US would follow AV3 and AV4 but not N3 and N4 (although this

prediction would itself be difficult to justify based on conditioning processes alone; see Dymond

and Roche, 2009). Because Augustson and Dougher (1997) did not measure ratings of the

likelihood of aversive stimuli during their study, there is no way of knowing whether

participants' expectancies and overt avoidance behaviour merged. Therefore, the present study

incorporated two ratings periods during avoidance learning and tests of indirect avoidance.

Second, accounts of the untrained control of avoidance behaviour based on sensory pre-

conditioning are necessarily limited by virtue of the involvement of only two directly paired

stimuli during the crucial training phase. Indeed, the avoidance responses and expectancies that

emerged in Declercq and De Houwer's (2009b) study for K and L were consistent with the

unidirectional nature of the pairing administered at the outset (i.e., A followed by K, B followed

by L; avoidance trained for A and B emerged for K and L, respectively). A convincing

demonstration of clinically relevant indirect avoidance would involve at least three bi-

directionally related equivalence relations separated by one intervening stimulus, similar to those

adopted by Augustson and Dougher (1997). Such a demonstration would speak to the forms of

avoidance that occur when, for instance, snake-phobic clients hear the word "snake" and other

unrelated stimuli, such as the word "reptile", pictures of snakes, names of different types of

snakes, a real snake, and places where snakes might be found that may occasion fear and avoidance. In effect, while the original CS for fear and avoidance may well have been directly conditioned, the generalization of fear and avoidance responses often occurs along verbal (i.e., symbolic) dimensions and can include a host of indirectly related, arbitrary stimuli.

Finally, extending the analysis of avoidance is essential in developing contemporary behavioural accounts of the emergence of clinical anxiety and avoidance (Field, 2006; Friman, Hayes, & Wilson, 1998; Rachman, 1977). Findings from a growing number of studies suggest that fear and anxiety often arise due to over-generalization of threat cues along non-arbitrary (physical) continua, such as emotional, perceptual and intensity continua (Dunsmoor, Mitroff, & LaBar, 2009; Dunsmoor, White, & LaBar, 2011; Lissek et al., 2008, 2010; Vervliet, Kindt, Vansteenwegen, & Hermans, 2010). However, if it can be shown that both avoidance and expectancy ratings may be shown to generalize along a non-physical, formal continuum, in the absence of any formal similarity or other direct associations between the stimuli, then such "symbolic generalization" may add a novel dimension to the understanding of generalized anxiety disorders.

The present study first established two equivalence relations each consisting of three stimuli (AV1-AV2-AV3 and N1-N2-N3) and then trained an avoidance (threat) response for one member of one of the relations (AV2) and a safety (non-avoidance) response for another (N2), before testing for indirect control of avoidance and ratings of the likelihood of aversive stimuli with presentations of AV3 and N3, respectively. Using this methodology, it is possible to probe for avoidance responses and US likelihood ratings using stimuli (AV3 and N3) that have never been directly paired with a CS+. Moreover, AV3 and N3 stimuli are related to the CSs only indirectly and along a purely symbolic (non-physical) stimulus dimension, thereby precluding the possibility of non-symbolic stimulus generalization.

Method

Participants

Twenty-one undergraduates were recruited from the student population of Swansea University and reimbursed with either £6 or partial course credit on completion of the study.

Apparatus

A computer program written in Visual Basic® 6.0 controlled all stimulus presentations and recorded all responses. Six nonsense words comprised the sample and comparison stimuli used during stimulus equivalence training and testing (i.e., JOM, CUG, VEK, BEH, PAF, ZID). Stimuli were presented in capitals, in uppercase bold size 24 Arial font. Two stimulus sets were constructed from these six stimuli and counterbalanced across participants.

Visual and auditory stimuli were selected from the *International Affective Picture System* (IAPS; Lang, Bradley, & Cuthbert, 2005) and the *International Affective Digitized Sounds* (IADS; Bradley & Lang, 1999) databases for use as aversive and nonaversive stimuli during the avoidance conditioning and derived transfer testing phases. A total of 20 photographs, 10 aversive (e.g., bodily mutilations) and 10 nonaversive (e.g., landscapes), and 10 aversive sounds (e.g., a female screaming) were selected[1]. The auditory stimuli were presented via headphones.

Procedure

On arrival at the laboratory, participants signed a consent form acknowledging the distasteful nature of some of the stimuli to be used during the experiment and indicating that they did not have a history of psychopathology. Participants were then seated comfortably at a table in front of a computer in a small experimental room.

Phase 1: Stimulus equivalence training and testing. During this phase, a delayed matching-to-sample (DMTS) procedure was used to train a series of conditional discriminations (AV1-AV2, AV1-AV3, N1-N2 & N1-N3) and test for the emergence of combined symmetry and

transitivity (i.e., stimulus equivalence) relations (AV2-AV3, N2-N3, AV3-AV2 & N3-N2; see

Figure 1). Participants were first given the following onscreen instructions, which were read

aloud by the experimenter:

> In a moment some words will appear on the screen. Look at the words at the top of the
>
> screen and then look at the two words at the bottom of the screen on the left and right.
>
> Choose one of the two words at the bottom of the screen by clicking on it. Sometimes the
>
> computer will give you feedback, and at other times it will not. However, you can get all
>
> of the tasks without feedback correct by carefully attending to the tasks with feedback.
>
> Press to continue.

On every trial, a nonsense word (AV1 or N1) first appeared in the top centre of the computer

screen (called the sample stimulus) for 1500 ms and was immediately followed by two further

nonsense words (AV2 and N2 or AV3 and N3) positioned in the bottom left and right corners of

the screen (called the comparison stimuli). No observing response was required to the sample,

and the comparisons remained on-screen until a response was made. Participants selected a

comparison by clicking on it with the computer mouse. When AV1 was presented, clicking on

the comparison stimulus AV2 produced the feedback, "Correct" in the centre of the screen, while

clicking on N2 produced the feedback "Wrong". When N1 was presented, clicking on the

comparison stimulus N2 produced the feedback, "Correct" in the centre of the screen, while

clicking on AV2 produced the feedback "Wrong". When AV1 was presented, clicking on the

comparison stimulus AV3 produced the feedback, "Correct" in the centre of the screen, while

clicking on N3 produced the feedback "Wrong". When N1 was presented, clicking on the

comparison stimulus N3 produced the feedback, "Correct" in the centre of the screen, while

clicking on AV3 produced the feedback "Wrong" (see Figure 1).

Feedback was displayed in size 14 Arial black font within a 4.5 x 2 cm square in the middle of the screen for 2 s, and was followed by an intertrial interval (ITI) of 2 s. A brief audible 'beep' also accompanied "Correct". All four tasks (AV1-AV2, AV1-AV3, N1-N2 & N1-N3) were presented in a block of 8 trials (each presented twice) in a pseudorandom order, with the constraint that the same task was not presented across more than two consecutive trials. Blocks were repeated until a participant made 8 consecutively correct responses.

On meeting the training criterion, a block of 16 trials were presented that tested for the emergence of combined symmetry and transitivity (i.e., stimulus equivalence) relations. Each of the four tasks (AV2-AV3, N2-N3, AV3-AV2 & N3-N2) was presented four times in the absence of feedback. When AV2 was presented, clicking on the comparison AV3 not N3, when N2 was presented, clicking on the comparison N3 not AV3, when AV3 was presented, clicking on the comparison AV2 not N2, and when N3 was presented, clicking on the comparison N2 not AV2, was predicted (Figure 1). Mastery criterion to infer the emergence of stimulus equivalence relations was set at 16 consecutive correct responses. If participants failed to achieve this, they were re-exposed to training and repeated testing until this criterion was met.

Phase 2: Threat and safety cue training. The purpose of this phase was to learn to avoid AV2 and to learn safety (non-avoidance) to N2 (see Figure 1). Participants were read the following on-screen instructions:

In a moment, you will be presented with some nonsense words, pictures and sounds. The pictures and sounds are from real life events and may be considered upsetting to some people. Pictures will be presented on the computer screen and sounds will be presented via headphones. Your task is to learn to cancel pictures and sounds before they are presented, by pressing the space bar. Later, you will be asked to make some ratings, by using a slider-scale, about the pictures and sounds.

Please follow the on-screen instructions and make your ratings as honestly as

possible. 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. When you are ready

to begin, press any key to continue.

Once participants had clicked the screen to proceed, a blank screen was displayed for

1700 ms. Next, either AV2 or N2 appeared in the centre of the screen for 5 s. If

participants pressed the space bar while either stimulus was present, then the screen

cleared and the words ''Picture Cancelled'' appeared for 2 s. If participants did not press

the space bar, the AV2 or N2 stimulus was followed by a 2 s interval, after which either a

600 x 800 pixel photograph and a sound (following AV2) or a blank screen was presented

for 2 s (following N2).

Aversive images and sounds followed all presentations of AV2 when the space bar

was not pressed (i.e., 100% contingency between non-avoidance and presentation of the

US). A blank screen followed all presentations of N2 when the space bar was not pressed

(i.e., 100% contingency between non-avoidance and absence of the US). The AV2 and N2

stimuli were presented in a pseudorandom order (i.e., no more than two consecutive

exposures to either) until participants made 6 consecutive avoidance responses during

presentations of AV2. If necessary, tasks were re-presented until participants

demonstrated conditioned avoidance according to this criterion.

On meeting the avoidance conditioning criteria, the first ratings period was introduced.

Participants were given four individual, randomly displayed questions about the likelihood of

pictures and sounds being presented both with and without the avoidance response during AV2

and N2 trials. The questions read as follows: "*How likely is it that the pictures and sounds will be*

presented if [AV2/N2] appears and you [do/do not] press the space bar? Please use the slider

scale below to rate how likely you think it is, where 1 is very unlikely and 10 is very likely."

Participants moved the slider scale with the computer mouse and confirmed their rating by

clicking on a button labeled "confirm [value chosen]". The ratings period ended once all four

questions were rated.

　　　　Phase 3: Inferred threat and safety cue testing. Phase 3 began immediately after Phase

2 with the onset of the following instructions:

　　　In a moment, you will be presented with nonsense words, pictures and sounds. The

　　　nonsense words will be the same as before, with some other nonsense words also presented

　　　that you will have seen before. Again, your task is to learn to cancel the pictures and sounds

　　　before they are presented. Also again, you will be asked to make some ratings, by using a

　　　slider-scale, about the pictures and sounds. Please follow the on-screen instructions and

　　　make your ratings as honestly as possible. It is important that you pay attention and

　　　concentrate on the screen at all times. When you are ready to begin, press any key to

　　　continue.

As before, once participants had pressed a key to proceed, a blank screen was displayed for 1700

ms, followed by a block of trials that presented AV2, AV3, N2, and N3 (see Figure 1). Stimuli

remained on screen for 5 s if no response occurred. If a participant pressed the space bar during

AV3 or N3, then the screen cleared and the words ''Picture Cancelled'' appeared for 2 s. Because

this was a test phase, not pressing the space bar during the presentation of AV3 and N3 was never

followed by an aversive image or sound (i.e., 0% contingency between AV3 and N3 and the

presence of the US). If a participant did not press the space bar during AV2 or N2, the same

contingencies were in place as in Phase 2. Five presentations each of AV3 and N3, and two

presentations each of AV2 and N2 were given, in a fourteen trial block. However, due to a

programming error, sixteen participants were given four presentations each of AV3 and N3, and

two presentations each of AV2 and N2, in a twelve-trial block. All trials were presented in a

pseudorandom order with the only constraint that no more than two consecutive trials of the same

type could occur.

After the test trials, the second ratings period was presented. As in Phase 2, participants

were given eight individual, randomly displayed questions about the likelihood of pictures and

sounds being presented both with and without the avoidance response during AV3, N3, AV2 and

N2 trials. Ratings were first obtained for situations describing AV3 and N3, followed by AV2

and N2. Following completion of the final rating, participants were thanked for their participation

and fully debriefed.

Dependent measures and analyses.

During *Phase 2* and *Phase 3*, the dependent measure was the percentage of trials with an

avoidance response to the learned threat cue, learned safety cue, inferred threat cue and inferred

safety cue. We evaluated the hypotheses that (a) the learned threat cue would prompt greater

avoidance than the safety cue during training and testing and (b) the inferred threat cue would

prompt greater avoidance than the inferred safety cue during testing. We hypothesized (a)

training would prompt higher ratings of the likelihood of aversive stimuli for *non-responding* to

the learned threat cue as compared to the learned safety cue and, similarly, (b), testing would

highlight higher ratings of the likelihood of aversive stimuli for *non-responding* to the inferred

threat cue as compared to the learned safety cue. Additionally, we hypothesized all learned and

inferred cues would evidence low ratings when engaging in avoidance. Thus, no differences were

anticipated between learned threat and safety or between inferred threat and safety. All

hypotheses were evaluated using separate paired *t*-tests corrected for multiple comparisons using

a *p* < .0045.

Results

During Phase 1, participants required a mean of 46.9 ($SD = 42.8$) training trials to meet criterion and a mean of 3.0 ($SD = 2.7$) cycles to pass the equivalence test.

Figure 2 shows the percentage of trials in which participants chose to avoid learned threat and safety cues during training (Phase 2) and inferred threat and safety cues during testing (Phase 3). In general and consistent with our predictions, avoidance behaviour was high to learned and tested threat cues and low to learned and tested safety cues. Results from Phase 2 showed that avoidance and safety cue training successfully produced a significantly high percentage of avoidance behaviour to the learned threat cue and considerably less avoidance behaviour to the learned safety cue, $t(20) = 2.083$, $p < .01$, (see Table 1). During Phase 3, this training effect was maintained such that there was significantly more avoidance behaviour to the learned threat cue than the safety cue, $t(20) = 8.041$, $p < .01$. The maintenance of the function of the learned cues provides important evidence that the testing context did not alter the threat and safety evoking properties of the cues. Results of testing showed the level of inferred avoidance was significantly greater for AV2 and AV3 than for N2 and N3 (see Table 1). Moreover, the difference in the levels of avoidance responses to the inferred threat and inferred safety cues, $t(20) = 4.933$, $p < .01$, was comparable to the difference observed between the learned threat and learned safety cues. Collectively, these findings illustrate how inferences about threat and safety can arise from verbal relational processes and motivate avoidance.

Figure 3 (bottom) shows the percentage of trials in which participants rated the likelihood of encountering the aversive stimulus when the avoidance response was made. In general, ratings were low and did not differ significantly when the avoidance response was made to the threat and safety cues after both training and testing. Figure 3 (top) shows the percentage of trials in which subjects rated the likelihood of encountering the aversive stimulus when the avoidance response

was not emitted. In general, ratings were high during training and testing to threat cues and low

during training and testing to safety cues. Moreover, ratings were consistent with patterns of

avoidance behaviour seen in Figure 2, which provides important convergent evidence for the

main effects. In training, ratings were significantly higher to the threat cue than the safety cue,

$t(20) = 3.823$, $p < .01$. Following training, this effect was maintained for the learned threat cue

and the learned safety cue, $t(20) = 1.330$, $p < 01$. Crucially, ratings during testing differed

significantly between the inferred threat cue and the inferred safety cue, $t(20) = 3.096$, $p < .01$,

while ratings between the learned and inferred threat cues and between learned and inferred

safety cues were not significantly different.

Discussion

The present experiment provided evidence of the indirect control of avoidance behaviour

and ratings in a laboratory model of inferred threat and safety learning. Following the emergence

of two equivalence relations each consisting of three stimuli (AV1-AV2-AV3 and N1-N2-N3),

threat-avoidance was trained for one member of one of the relations (AV2) and safety (non-

avoidance) for another (N2). Then, inferred threat-avoidance and safety behaviour and US

expectancies were tested with presentations of all remaining stimuli. Findings showed that threat

(avoidance) and safety cue (non-avoidance) training produced a significantly high percentage of

avoidance to the learned threat cue and considerably less avoidance to the learned safety cue, and

that ratings were high during training and testing to threat cues and low to safety cues. The

magnitude of the significant differences between learned threat and safety cues were preserved

from training to testing. Importantly, following testing, ratings showed there was a high

likelihood of aversive stimuli without avoidance responding following the inferred threat cue and

a low likelihood without avoidance responding following the inferred safety cue. The high ratings

of the likelihood of contacting the aversive stimulus for the inferred threat cue is significant

because it highlights the formation of a threat belief that stems from symbolic stimulus equivalence relations and not from directly contacting the aversive stimulus (experience). Collectively, these findings are the first to show how inferences about threat and safety can arise from relational learning processes and generate ratings of the likelihood of aversive events (Augustson & Dougher, 1997; Dymond et al., 2007, 2008; Dymond & Roche, 2009; Gannon et al., in press; Roche et al., 2008).

The present findings are consistent with the expectancy model of avoidance (Declercq & De Houwer, 2009b, 2011; Lovibond, 2006). Participants made the avoidance response in the presence of AV3 and rated it as more likely to be followed by the US than N3, with no direct training. Moreover, participants did not make the avoidance response in the presence of N3 and rated it as less likely to be followed by the US than AV3. These novel threat and safety behaviours emerged by virtue of the participation of AV2 and AV3 and N2 and N3 in stimulus equivalence relations such that actions and expectancies that occurred with one set of stimuli readily transferred to the remaining, indirectly related stimuli. The findings from threat and safety cue training indicated that participants learned that AV2 would be followed by the US and that making the avoidance response resulted in the US being omitted. Lovibond's (2006) expectancy model predicts that the resulting propositional knowledge (Mitchell, De Houwer, & Lovibond, 2009) leads to the inferences that in the absence of the avoidance response, AV2 will be followed by the US and that, in the presence of the avoidance response, AV2 will not be followed by the US. Presumably, the emergence of avoidance behaviour in the presence of AV3 resulted from the expectancy that in the absence of the avoidance response, AV3 will be followed by the US, and that in the presence of the avoidance response, AV3 will not be followed by the US. In this way, it is possible to extend the predictions of the revised expectancy model of avoidance to the present findings involving stimulus equivalence relations. However, the present findings raise

several important challenges for the expectancy account of avoidance. First, avoidance was never directly learned in the presence of AV3, yet reliable avoidance behaviour still occurred. US ratings also showed modulation by the presence or absence of the avoidance response during AV3 presentations. For Lovibond's (2006) model to explain these findings it is necessary to assume that the indirect threat and safety test phase established a similar context to that of the learning phase which generated expectancies concerning the likely omission of the US in the presence of the avoidance response and the non-omission of the US in the absence of the avoidance response to AV3. Thus, while a revised form of the expectancy model may account for the present findings, further research is needed to determine the critical features of the testing context that occasion indirect avoidance and ratings of aversive outcomes.

Second, to date, studies examining indirect control of avoidance have involved direct pairings of stimuli (Declercq & De Houwer, 2009b; Lovibond et al., 2007, 2009; see also, Dunsmoor et al., 2009, 2011). In the present study, indirect avoidance was observed with AV3 that had not been directly paired with the learned threat cue, AV2. The DMTS procedure used in Phase 1 ensured that during testing sample stimulus offset (i.e., AV2) was immediately followed by onset of the comparisons (i.e., AV3 and N3; see Figure 1). The fact that the sample and predicted comparison(s) never appeared simultaneously onscreen makes it highly unlikely that avoidance emerged through sensory pre-conditioning, second-order conditioning or stimulus compounding processes because the stimuli that occasioned avoidance, AV2 and AV3, were never directly paired (Hall, 1996; Rehfeldt & Hayes, 1998; Smeets & Barnes-Holmes, 2003). The low number of exposures required to pass the stimulus equivalence test (*M:* 3) further indicates it was unlikely that unspecified forms of inadvertent, adventitious feedback could have influenced responding. In conclusion, while the present findings may be explained in accordance with the expectancy model of avoidance, explaining the emergence of avoidance *solely* in terms of

associative learning processes is not parsimonious and necessitates a consideration of the role of verbal relational processes in the acquisition of indirect avoidance (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Dougher, Hamilton, Fink, & Harrington, 2007; Dymond & Roche, 2009; Hayes & Hayes, 1992; Smyth, Barnes-Holmes, & Forsyth, 2006).

Thirdly, Lovibond's (2006) expectancy model may have heuristic value, but it must be remembered that the causal relationship between expectancy and avoidance remains inferential. Specifically, while expectancies may indeed mediate avoidance responses in the current study, it is equally likely that the equivalence relations established by the experimenters functioned as the causal mechanism for *both* the expectancies and the avoidance responses. Thus, according to the current study outcome, expectancy itself can be explained in terms of the relational contingencies instituted by the experimenters. Such a view is more parsimonious (i.e., has fewer assumptions) than an expectancy-based account, and is a popular, functional view among those in derived relational responding research (e.g., Hayes, Barnes-Holmes & Roche, 2001; Hayes & Hayes, 1992; Sidman, 1994; Smyth, Barnes-Holmes, & Barnes-Holmes, 2008; see also, De Houwer, 2011; Hughes, Barnes-Holmes, & De Houwer, in press, for related arguments). In simple terms, an expectancy may be viewed as itself consisting of a discriminated stimulus relation and so constitutes the outcome of a relational learning process, rather than the immediate cause of relational learning effects, such as indirect avoidance (Dymond & Roche, 2009; Dymond et al., 2007, 2008; Friman et al., 1998; Roche et al., 2008). In effect, from our perspective, relational learning functions as a "third variable" that can explain the emergence of both expectancies *and* avoidance responses. In this way, expectancy and avoidance represent instances of relational responding. This functional, behaviour-analytic approach may be distinguished from the cognitive/mechanistic perspective that emphasizes inferences and expectancies as mediational

constructs (De Houwer, 2011; Dymond & Roche, 2009; Friman et al., 1998; Hayes & Brownstein, 1986).

Our use of the "picture cancelled" feedback screens following avoidance in the presence of all stimuli during threat and safety cue training and testing warrants comment. Despite this common contingency, which may resemble more than just the absence of the scheduled US, participants readily met the avoidance criterion and made the appropriate response to AV2 (and AV3) and withheld responding to N2 and N3. Clearly, this feedback was inaccurate following N2 and N3 presentations, since the US was never scheduled to follow these stimuli regardless of responding. Participants could, therefore, have merely pressed the space bar on all trials and avoided all US presentations. The data indicate that this pattern of behaviour did not emerge; avoidance occurred on a mean of only 4.3% of learned safety cue trials (Figure 2). There may have been several reasons for this. First, we employed a stringent avoidance-conditioning criterion similar to that adopted by previous studies (Dymond et al., 2007, 2008) in which a minimum of six consecutive avoidance responses to AV2 were required. This emphasis on maintenance of response rate, set within a predetermined maximum number of avoidance trial exposures, maintained avoidance under free-operant conditions (Higgins & Morris, 1984). This was necessary in order to present multiple stimulus probes during the critical testing phase. Second, the instructions emphasized that participants should learn to press the space bar in the presence of some stimuli but not others. This may have prompted participants to conserve responding during N2 (and N3) trials. Finally, had the feedback screens been sufficient to evoke avoidance of all stimuli no participants would have met the conditioning criterion. Clearly, future research on the role of feedback in acquiring and maintaining free-operant avoidance behavior is warranted.

Conditioning models have long been challenged to explain the complexity and diversity

of clinical problems, particularly the necessity of a prior traumatic conditioning history with the

feared object or event in order to explain the acquisition of anxiety disorders (Rachman, 1977).

Contemporary theorists postulate vicarious or alternative pathways to fear and anxiety (e.g.,

Field, 2006; Mineka & Oehlberg, 2008; Muris & Field, 2011) that appear to be based on direct

conditioning processes (e.g., Bouton, Mineka, & Barlow, 2001; Mineka & Oehlberg, 2008).

Others have considered the top-down influence of high-level cognitive processes (Lovibond &

Shanks, 2002; Lovibond et al., 2009; Mitchell et al., 2009). However, one important contribution

of the current findings is the demonstration that inferred avoidance and attendant ratings were

generated for stimuli using a relational learning procedure. In effect, avoidance and ratings of the

likelihood of aversive stimuli showed symbolic generalization along a non-formal continuum

from AV2 and N2 to AV3 and N3, respectively, in the absence of any formal similarity or other

direct associations between the stimuli. A traditional conceptualization of stimulus generalization

may account for fear responses spreading to stimuli that formally resemble the originally trained

CS (Blough, 1975; Bouton et al., 2001; Mineka & Oehlberg, 2008). However, findings from the

literature on fear generalization cannot be easily applied to the outcomes of the current study, as

it did not involve the indirect acquisition of threat related responses across physically similar

stimuli. Instead, our findings show that the generalization of avoidance responses can occur along

verbal (i.e., symbolic) pathways, which has implications for the treatment of anxiety disorders

and phobias. Lissek et al. (2010) suggested that panic disorder patients may be trained to

discriminate the boundaries of the realistic threat stimulus class from those stimuli only distantly

related to threat. Clinically, however, the class of potential threat cues comprises a large and

diverse range of symbolically related stimuli (e.g., the spoken word "snake" bears no formal

resemblance to a real snake). In other words, the threat function of a stimulus can expand rapidly

through symbolic generalization. In cases such as this, treatment methods drawn from "third-wave" behaviour therapies, such as Acceptance and Commitment Therapy (Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Hayes, Villate, Levin, & Hildebrandt, 2011), may prove promising and warrant further attention (e.g., defusion exercises; see Masuda, Feinstein, Wendell, & Sheehan, 2010).

In conclusion, much remains to be investigated about the basic behavioural processes of human avoidance, its underlying neurocircuitry (see Schlund et al., 2010; Schlund & Cataldo, 2010), and the implications for therapy. Further laboratory research is warranted on, for instance, stimulus relations such as more than/less than (Dougher et al., 2007; Munnelly, Dymond, & Hinton, 2010) in generating increased levels of fear-relevant avoidance and on differentiating potential outcomes in inferred threat and safety learning with sub-clinical and clinically anxious groups. The present findings may have potential as a novel means of investigating the neurobehavioural determinants of symbolic generalization in human avoidance learning and how neuropathology may contribute to excessive forms of avoidance seen in clinical disorders.

Acknowledgments

References

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

Blough, D. S. (1975). Steady state data and quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 3–21.

Bouton, M. E., Mineka, S., & Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorder. *Psychological Review*, *108*, 4–32.

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings* (Tech. Rep. No. B-2). Gainesville, FL: University of Florida, Center for Research in Psychophysiology.

Declercq, M., & De Houwer, J. (2008). On the role of US expectancies in avoidance behaviour. *Psychonomic Bulletin & Review*, *15*, 99-102.

Declercq, M., & De Houwer, J. (2009a). Evidence for a hierarchical structure underlying avoidance behaviour. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 123–128.

Declercq, M., & De Houwer, J. (2009b). Transfer of avoidance responding to a sensory preconditioned cue: Evidence for the role of S-S and R-S knowledge in avoidance learning. *Learning and Motivation*, *40*, 197–208.

Declercq, M., & De Houwer, J. (2011). Evidence against an occasion setting account of avoidance learning. *Learning and Motivation*, *42*, 46–52.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*, 202-209.

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.

Dougher, M.J., Hamilton, D.A., Fink, B.C. & Harrington, J. (2007). Transformation of the

discriminative and eliciting functions of generalized relational stimuli. *Journal of the

Experimental Analysis of Behavior*, *88*, 179-197.

Dunsmoor, J.E., Mitroff, S.R., & LaBar, K.S., (2009). Generalization of conditioned fear along a

dimension of increasing fear intensity. *Learning & Memory*, *16*, 460–469.

Dunsmoor, J. E., White, A. J., & LaBar, K. S. (2011). Conceptual similarity promotes

generalization of higher-order learning. *Learning & Memory*, *18*, 156-60.

Dymond, S., & Roche, B. (2009). A contemporary behaviour analysis of anxiety and avoidance.

The Behavior Analyst, 32, 7-28.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). Transformation of

avoidance response functions in accordance with the relational frames of same and

opposite. *Journal of the Experimental Analysis of Behavior*, *88*, 249-262.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2008). Derived avoidance

learning: Transformation of avoidance response functions in accordance with the relational

frames of same and opposite. *The Psychological Record*, *58*, 271-288.

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and

treatment of phobias? *Clinical Psychology Review*, *26*, 857–875.

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behaviour analysts should study

emotion: The example of anxiety. *Journal of Applied Behavior Analysis*, *31*, 137–156.

Gannon, S., Roche, B., Kanter, J., Forsyth, J. P., & Linehan, C. (in press). A derived relations analysis of approach-avoidance conflict: Implications for the behavioural analysis of human anxiety. *The Psychological Record.*

Hall, G., (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Learning & Behavior*, *24*, 233-255.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behaviour-behaviour relations, and a behaviour-analytic view of the purposes of science. *The Behavior Analyst, 9*, 175-190.

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behaviour analysis. *American Psychologist*, *47*, 1383–1395.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes, and outcomes. *Behaviour Research and Therapy*, *44*, 1–25.

Hayes, S. C., Villate, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioural and cognitive therapies. *Annual Review of Clinical Psychology*, *7*, 141-168.

Higgins, S. T., & Morris, E. K. (1984). Generality of free-operant avoidance conditioning to human behavior. *Psychological Bulletin*, *96*(2), 247-272.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (in press). The dominance of associative theorizing in implicit attitude research: Propositional and behavioural alternatives. *The Psychological Record*.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Tech. Rep. No. A-6). Gainesville, FL: University of Florida, Center for Research in Psychophysiology.

Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon, C., (2008). Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. *Behaviour Research and Therapy*, *46*, 678–687.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C., (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry*, *167*, 47–55.

Lovibond, P. F. (2006). Fear and avoidance: An integrated expectancy model. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: Basic science to clinical application* (pp. 117–132). Washington, DC: American Psychological Association.

Lovibond, P. F., Saunders, J. C., Weidemann, G., & Mitchell, C. J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *Quarterly Journal of Experimental Psychology*, *61*, 1199–1216.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 3-6.

Lovibond, P. F., Mitchell, C. J., Minard, E., Brady, A., & Menzies, R. G. (2009). Safety behaviours preserve threat beliefs: Protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy, 47*, 716–720.

Ly, V., & Roelofs, K. (2009). Social anxiety and cognitive expectancy of aversive outcome in avoidance conditioning. *Behaviour Research and Therapy*, *47*, 840–847.

Masuda, A., Feinstein, A. B., Wendell, J. W., & Sheehan, S. T. (2010). Cognitive defusion versus thought distraction: A clinical rationale, training, and experiential exercise in altering psychological impacts of negative self-referential thoughts. *Behavior Modification*, *34*, 520-538.

Mineka, S., & Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychologica*, *127*, 567-580.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183–246.

Munnelly, A., Dymond, S., & Hinton, E. C. (2010). Relational reasoning with derived comparative relations: A novel model of transitive inference. *Behavioural Processes, 85*, 8-17.

Muris, P., & Field, A. P. (2011). The role of verbal threat information in the development of childhood fear. ''Beware the Jabberwock!'' *Clinical Child & Family Psychology Review,* 13, 129–150.

Rachman, S. J. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy*, *15*, 375–387.

Rachman, S., Radomsky, A. S., & Shafran, R. (2008). Safety behaviour: A reconsideration. *Behaviour Research and Therapy*, *46*, 163-173.

Rehfeldt, R. A., & Hayes, L. J. (1998). The operant-respondent distinction revisited: Toward an understanding of stimulus equivalence. *The Psychological Record*, *48*, 187-210.

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.

Salkovskis, P. M. (1991). The importance of behaviour in the maintenance of anxiety and panic:

A cognitive account. *Behavioural Psychotherapy*, *19*, 6–19.

Salkovskis, P. M., Clark, D. M., Hackmann, A., Wells, A., & Gelder, M. G. (1999). An

experimental investigation of the role of safety-seeking behaviours in the maintenance of

panic disorder with agoraphobia. *Behaviour Research and Therapy*, *37*, 559–574.

Schlund, M.W., & Cataldo, M. F. (2010). Amygdala involvement in human avoidance, escape

and approach behaviour. *NeuroImage 53*, 769-776.

Schlund, M. W., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Cataldo, M. F., Forbes, E. E., Dahl,

R. E., & Ryan N. D. (2010). Nothing to fear? Neural systems supporting avoidance

behaviour in healthy youths. *NeuroImage, 52*, 710-719.

Sidman, M., (1994). *Equivalence relations and behavior: A research story*. Boston: Authors'

Cooperative.

Smeets, P. M., & Barnes-Holmes, D. (2003). Children's emergent preferences for soft drinks:

Stimulus-equivalence and transfer. *Journal of Economic Psychology*, *24*, 603–661.

Smyth, S., Barnes-Holmes, D., & Barnes-Holmes, Y. (2008). Acquired equivalence in human

discrimination learning: The role of propositional knowledge. *Journal of Experimental

Psychology: Animal Behaviour Processes, 34*, 167-177.

Smyth, S., Barnes-Holmes, D. & Forsyth, J. P. (2006). Derived transfer of self-reported arousal

functions. *Journal of the Experimental Analysis of Behavior*, *85*, 223-246.

Vervliet, B., Kindt, K., Vansteenwegen, D., & Hermans, D. (2010). Fear generalization in

humans: Impact of prior non-fearful experiences. *Behaviour Research and Therapy*, 48,

1078-1084.

Figure Captions

Figure 1: Overview of the experimental procedure. Arrows indicate trained relations and responses, and '?' indicate predicted, tested responses. See text for details.

Figure 2: Percentage avoidance to learned threat and safety cues during training and learned and indirect/inferred threat and safety cues during testing. Error bars represent standard error.

Figure 3: Participants' ratings of the likelihood of encountering the aversive stimulus when the avoidance response was (lower panel) and was not made (upper panel) during threat and safety training and testing phases. Error bars represent standard error.

Footnote

[1] IAPS and IADS identifiers: (Pictures) #3000, #3010, #3030, #3051, #3060, #3061, #3062,

#3063, #3064, #3068; (Sounds) #276, #277, #278, #279, #285, #286, #290, #292, #380, #423.

- Symbolic generalization of avoidance may underlie the etiology and maintenance of anxiety.

- We demonstrate avoidance of directly learned and indirectly related stimuli.

- Findings have implications for behavioural accounts of anxiety.

Table 1. Mean (and standard deviation) percent avoidance responses and ratings during threat and safety training and testing phases. Also shown is Cohen's *d* for threat-safety comparisons.

Measure	Phase	Stimulus	Mean	SD	Threat-Safety *d'*
Percent Avoidance	2: Training	Learned Threat	82.54	7.90	8.86
		Learned Safety	4.40	9.66	
	3: Testing	Learned Threat	95.24	15.04	8.96
		Learned Safety	0.00	0.00	
		Inferred Threat	89.76	21.59	3.53
		Inferred Safety	9.05	24.06	
Rating: No Avoidance	2: Training	Learned Threat	9.10	2.70	2.13
		Learned Safety	2.71	3.27	
	3: Testing	Learned Threat	9.38	1.99	4.00
		Learned Safety	1.48	1.97	
		Inferred Threat	8.48	2.66	3.65
		Inferred Safety	1.24	0.89	
Rating: Avoidance	2: Training	Learned Threat	3.52	4.09	0.52
		Learned Safety	1.86	1.85	
	3: Testing	Learned Threat	2.00	2.68	-0.39
		Learned Safety	3.14	3.12	
		Inferred Threat	2.81	3.37	-0.15
		Inferred Safety	3.29	3.16	

Phase 1: Equivalence Training & Testing

Avoidance cues

Train

AV1 → AV2 / N2

AV1 → AV3 / N3

Test

AV2 → AV3 / N3

AV3 → AV2 / N2

Neutral cues

Train

N1 → N2 / AV2

N1 → N3 / AV3

Test

N2 → N3 / AV3

N3 → N2 / AV2

Phase 2: Threat & Safety Cue Learning

AV2 → Avoidance / Non-avoidance

N2 → Avoidance / Non-avoidance

Phase 3: Inferred Threat & Safety Cue Testing

AV2 → Avoidance / Non-avoidance

AV3 → Avoidance? / Non-avoidance?

Does avoidance occur to AV3?

N2 → Avoidance / Non-avoidance

N3 → Avoidance? / Non-avoidance?

Does safety (non-avoidance) occur to N3?

Expectancies: No Avoidance

Training | Testing

Very Likely / Very Unlikely

Learned Threat AV2 | Learned Safety N2 | Learned Threat AV2 | Inferred Threat AV3 | Learned Safety N2 | Inferred Safety N3

Expectancies: Avoidance

Training | Testing

Very Likely / Very Unlikely

Learned Threat AV2 | Learned Safety N2 | Learned Threat AV2 | Inferred Threat AV3 | Learned Safety N2 | Inferred Safety N3