One-Trial Overshadowing:

Evidence For Fast Specific Fear Learning in Humans

Kim Haesen^{1, 2}, Tom Beckers^{1, 2}, Frank Baeyens^{1, 2}, and Bram Vervliet^{1, 2, 3, 4}

¹Center for Learning Psychology and Experimental Psychopathology, KU Leuven, Leuven

²Center for Excellence on Generalization in Health and Psychopathology, KU Leuven,
Leuven

³Department of Psychiatry, Harvard Medical School, Boston

⁴Department of Psychiatry, Massachusetts General Hospital, Boston

KU Leuven, Tiensestraat 102, 3000 Belgium

kim.haesen@kuleuven.be, tom.beckers@kuleuven.be, frank.baeyens@kuleuven.be,
bram.vervliet@kuleuven.be

Corresponding author: Bram Vervliet

tel:+32 16 32 61 17

Abstract

Adaptive defensive actions necessitate a fear learning system that is both fast and specific. Fast learning serves to minimize the number of threat confrontations, while specific learning ensures that the acquired fears are tied to threat-relevant cues only. In Pavlovian fear conditioning, fear acquisition is typically studied via *repetitive* pairings of a *single* cue with an aversive experience, which is not optimal for the examination of fast specific fear learning. In this study, we adopted the one-trial overshadowing procedure from basic learning research, in which a combination of two visual cues is presented once and paired with an aversive electrical stimulation. Using on-line shock expectancy ratings, skin conductance reactivity and startle reflex modulation as indices of fear learning, we found evidence of strong fear after a single conditioning trial (fast learning) as well as attenuated fear responding when only half of the trained stimulus combination was presented (specific learning). Moreover, specificity of fear responding tended to correlate with levels of state and trait anxiety. These results suggest that one-trial overshadowing can be used as a model to study fast specific fear learning in humans and individual differences therein.

Keywords: Anxiety disorders, Overshadowing, Fear Conditioning, Generalization, State Anxiety, Trait Anxiety

1. Introduction

Fear is the ensemble of emotional reactions that is triggered when danger is imminent. It serves to motivate defensive actions that cope with the upcoming dangerous event. While some fears are genetically pre-programmed, most fears are acquired through Pavlovian conditioning (Davey, 1992; Lissek et al., 2005). This consists of learning experiences in which dangerous encounters (unconditional stimulus, US) become associated with preceding innocuous cues (conditional stimulus, CS) that henceforth signal the US and elicit fear (conditional reaction). In the natural environment, fear learning is adaptive when it is *fast* (in order to minimize confrontations with dangerous events) and *specific* to the actual situation that carries the danger potential.

Fast-yet-specific fear learning is challenged by the fact that danger situations are often complex in nature and involve multiple preceding cues (i.e., potential CSs). Failures in fast specific learning under these circumstances may contribute to fear overgeneralization patterns as typically seen in anxiety-related disorders. Post-traumatic stress disorder, for example, is characterized by psychological and physiological distress during exposures to subsets of cues that resemble aspects of the traumatic event (American Psychiatric Association, 2013; Ehlers et al., 2002; Ehlers & Clark, 2000). However, the challenge of fast-specific fear learning is not readily captured in the prototypical fear conditioning procedure where a *single* stimulus (CS) is paired *multiple* times with an aversive stimulus (US). The purpose of the current study was to develop a pre-clinical human fear conditioning protocol for the study of fast-specific fear learning, given its relevance for adaptive and clinical forms of anxiety.

The standard fear conditioning procedure has been adjusted before to study either fast or specific learning. First, studies showed that humans (Öhman, Eriksson, & Olofsson, 1975) and non-human animals (Fanselow, 1990; Jarvik & Essman, 1960) can learn to fear simple neutral stimuli and even more complex contexts following a *single* learning trial (fast

learning). Second, *specificity* has been studied via tests of fear generalization between simple stimuli that vary on a single perceptual dimension (Dymond et al., 2015). For example, fear conditioning to a circle is followed by generalization tests with a range of circles of varying sizes (Lissek et al., 2008). This typically results in an orderly descending fear gradient as a function of perceptual similarity. Context generalization research extended this research to more complex situations that simultaneously differ on multiple perceptual dimensions (Rudy & O'Reilly, 1999; Rudy & O'Reilly, 2001). Hence, fear generalization research has focused mainly on the effects of gradual changes in conditioned stimuli and contexts. Much less is known, however, about the degree to which fear conditioned to a stimulus complex generalizes to its constituent stimuli when encountered separately. This lacuna is surprising; because conditions like PTSD are markedly characterized by inflated fear reactions to isolated trauma cues *as if* the whole traumatic situation is once again present. In sum, while procedures have been developed to examine either fast or specific fear learning, little is known about the combined challenge of fast-specific fear learning in complex stimulus situations.

Basic associative learning research in non-human animals has provided a simple, yet elegant procedure that lends itself perfectly to the study of fast-yet-specific learning in complex stimulus situations. One-trial overshadowing (OTO, a variant of the multiple-trial overshadowing procedure; Kaufman & Bolles, 1981; Lanzetta & Orr, 1980, 1981; Mackintosh, 1976; Pavlov, 1927) refers to a procedure in which a combination of two CSs is presented once, followed by the US. Subsequently, the CSs are presented separately and the degree of conditioned responding is measured. Under these conditions, non-human animals show lower conditioned responding compared to a CS that was individually paired with the US (Cole, Oberling, & Miller, 1999; James & Wanger, 1980; Kaye, Gambini, & Mackintosh, 1988; McNally & Westbrook, 2003; Mackintosh, 1971; Mackintosh & Reese, 1979; but see

Zelikowsky & Fanselow, 2010). According to standard associative learning theory, this reduction reflects *configural processing* of the CS complex during the conditioning trial (see Kaye, Gambini, & Mackintosh, 2007). With stronger configural processing, the fear association is thought to accrue to the stimulus complex as a whole, rather than to its constituent stimuli. Separated presentations of these stimuli will therefore fail to activate the fear association and trigger little fear (the OTO-effect). According to non-associative theories, on the other hand, the fear decrease may primarily reflect the *uncertainty/ambiguity* that surrounds the first confrontation with part of a conditioned stimulus compound (probabilistic/propositional theories of conditioning, De Houwer, Vandorpe, & Beckers, 2005; Mitchell, De Houwer, & Lovibond, 2009). Interestingly, anxiety-related conditions like PTSD have been linked both to configural processing deficits (e.g., Gilbertson et al., 2007) and threat biases during situations of uncertainty/ambiguity (e.g., Beckers et al., 2013; Lissek, Pine, & Grillon, 2006). These deficits/biases are expected to impair OTO. First, configural processing deficits may lead to the individual elements present during conditioning all acquiring an association with the US, resulting in excessive fear responding in OTO (see General Discussion). Second, threat biases would lead to an over-estimation of the threat value of the uncertainty/ambiguity surrounding OTO and trigger aberrant fear responding as well. Thus, OTO is both theoretically and clinically relevant to the study of anxiety.

The present study was set up to adapt and validate OTO as a model for fast-yet-specific fear learning in humans. A compound CS (AX) was paired with an aversive electrical stimulation once, followed by test presentations of one of the CSs (X). In addition, a single CS (B) was paired with the stimulation once, while another single CS (C) was not. Fast learning would be reflected by stronger fear reactions to B than to C ($B > C$). Selective learning would be reflected by lower fear reactions to X than to B ($B > X$). We also assessed state and

trait levels of anxiety to explore influences of anxiety on fast-selective fear learning in humans.

2. Experiment 1

2.1 Material and methods

2.1.1 Participants

Twenty psychology students (nine women) with a mean age of 23.4 (SD = 6.38) participated in experiment 1 in return for course credits. All participants gave informed consent and were aware that they could abort the experiment at any time.

2.1.2 Apparatus

Conditioned stimuli

Four geometrical shapes (trapezium, diamond, hexagon, cross) served as conditional stimuli and were presented on a computer screen (Dell LCD monitor, type 1708 FPc). These shapes were grey with a black border and presented in a white frame. Stimuli were presented against a black background screen.

Unconditioned stimuli

A 2-ms electrocutaneous stimulus administered to the wrist of the dominant hand served as a US. A Digitimer DS71 constant current stimulator (Hertfordshire, UK) administered the stimulus via a pair of V91-01-8-mm reusable Bilaney Ag/AgCL electrodes. Electrodes were filled with K-Y Jelly.

2.1.3 Measures

Questionnaires

Participants completed the Trait Anxiety inventory (STAI-T; Spielberger, 1983; Dutch translation by van der Ploeg, 2000).

Skin conductance reactivity

Electrodermal activity was recorded using a skin conductance coupler manufactured by Coulbourn Instruments (model V71-23, Allentown, PA). A constant voltage of 0.5 V was applied across a pair of 8-mm Ag/AgCl electrodes. These electrodes were attached to the palm of the non-dominant hand. A Labmaster DMA 12 bit analog-to-digital converter (Scientific Solutions, Solon, Ohio) digitized the skin conductance signal at 10 Hz from 2 s prior to CS onset until 6 s after CS offset.

Shock-expectancy

To measure trial-by-trial subjective shock expectancy ratings, an eleven-point scale was used. This scale ranged from 0 to 10 and was labeled: “certainly no shock” (0), “maybe” (5), “certain shock” (10). A left mouse click on the scale registered the corresponding position for that trial.

2.1.4 Procedure

After participants gave their informed consent, they completed the STAI-T questionnaire. Next, electrodes were fitted and the shock intensity was set to a level that was determined “definitely uncomfortable, but not painful” through a standard shock work-up procedure. Subsequently, participants were instructed that pictures of geometrical shapes would appear on the computer screen and that some of these shapes could be followed by a shock. It was further explained that the participant’s task was to predict the occurrence of the shock. Next, participants were instructed how to use the expectancy ratings scale.

The experiment consisted of two phases (see Table 1 and Figure 1). During acquisition two stimuli (A and X) were presented in compound (AX) and followed by an electrical shock. Two additional stimuli (B and C) were each presented individually. Stimulus B was followed by shock, stimulus C was not. Each stimulus or stimulus compound was presented only once. The geometrical shapes serving as X, B and C were counterbalanced as close as possible: in the final sample each stimulus combination (six combinations in total) appeared two or three

times. Stimuli were presented in random order. The last acquisition trial was followed by a 20-s interval during which no stimuli were presented and the screen remained black. The test phase was divided in three blocks. In each block all stimuli were presented once, individually. In the first block, the non-reinforced stimulus C was always presented first. Next, stimulus B and X were presented. The order in which these two stimuli were presented was counterbalanced. Stimulus A was always tested at the end of block 1. Stimulus presentation was randomized in the following two blocks.

Each CS was presented for eight seconds, with on average 14 s (range 12-16 s) between two CS presentations (from CS offset to CS onset). The scale appeared at the bottom of the screen 0.5 s following CS-onset. Participants moved around a red dot across the scale using the computer mouse and confirmed their rating with a left-mouse click after which the scale disappeared from the computer screen. During the training phase stimulus AX and B were followed by an electrical shock 7.5 seconds after stimulus onset. The stimulus sequence, stimulus presentation, ITI, and response registration was controlled by Affect 4.0 software (Spruyt et al., 2002).

2.2 Results

2.2.1 Data reduction

Due to a programming error, four participants had to be excluded from all analyses. Following visual inspection of SCR data, another participant was excluded from analysis of SCR. One participant did not complete the STAI-T questionnaire.

Skin conductance responses were extracted with the PSPHA software for analysing psychophysiological data (De Clerck, Verschuere, De Vlieger, & Crombez, 2006). Skin conductance response amplitudes to the CS were calculated by determining the peak of the skin conductance level between CS onset and 7500 ms after CS onset and subtracting the baseline SCR (calculated as the average skin conductance level during 2000 ms prior to CS

onset). Resulting SCR amplitudes were range-corrected by dividing the amplitudes by maximum SCR amplitude to the US (calculated by determining the peak of the skin conductance level between 1500 ms and 6500 ms following the shock minus the baseline SCR). The square root of these range-corrected amplitudes was used for statistical analysis.

3.2.2 Shock expectancy ratings

Single conditioning trial

The left panel of Figure 2A suggests that on the first and only training trial, prior to any conditioning, shock expectancies were equally strong for stimuli B, C and the AX compound. This was confirmed by a Repeated Measures (RM) ANOVA with one within-subjects factor (Stimulus, three levels), $F(1.34, 18.76) < 1$.

Test

The right panel of Figure 2A suggests higher shock expectancy ratings for B than for C and X on the first test trial. This was confirmed by a RM ANOVA with one within-subjects factor (Stimulus; three levels), which revealed a main effect of Stimulus, $F(2, 30) = 20.06, p < .001, \eta^2_{part} = .57$. Planned comparisons revealed a difference in shock expectancy ratings between B and C, $F(1, 15) = 31.65, p < .001, \eta^2_{part} = .68$ (one-trial learning), and B and X, $F(1, 15) = 15.76, p = .001, \eta^2_{part} = .21$ (one-trial selective learning).

3.2.3 Skin Conductance responses

Single conditioning trial

The left panel of Figure 2B suggests equally strong SCRs for AX, B and C on the first and only training trial. This was confirmed by a RM ANOVA with one within-subjects factor (Stimulus, 3 levels), which revealed no main effect of Stimulus, $F(2, 28) < 1$.

Test

The right panel of Figure 2B suggests higher SCRs for B than for C and X on the first test trial. This was confirmed by a RM ANOVA with one within-subjects factor (Stimulus, 3

levels) which showed a marginally significant main effect of Stimulus, $F(2, 28) = 2.78, p = .08, \eta^2_{part} = .166$. Planned comparisons revealed significantly higher SCRs for B than X, $F(1, 14) = 5.74, p = .03, \eta^2_{part} = .29$ (one-trial selective learning), but no difference in SCR between B and C, $F(1, 14) < 1$ (one-trial learning).

3.2.4 Questionnaires

To further explore the relation between selective learning and anxiety, correlations were calculated between STAI-T on the one hand and conditioned fear to X and a selective fear index on the other hand. The selective fear index was calculated by subtracting conditioned responding to X from conditioned responding to B on the first test trial. Because none of the questionnaire-data were distributed normally, non-parametric Spearman Rho correlations were calculated. Correlations were tested two-sided.

STAI-T

STAI-T scores showed a marginally significant negative correlation with the SCR selective learning index, $\rho = -.48, p = .08$, indicating that participants with higher trait anxiety showed less selective learning. No significant correlation between STAI-T scores and SCR on the first X test-trial were observed, $\rho = .37, p = .19$. STAI-T scores did not correlate with shock expectancy ratings on the first X test trial or the selective learning index of shock expectancies, all $\rho s < .262$, all $p s > .35$.

3.3 Discussion

Experiment 1 established the one-trial overshadowing effect in human fear conditioning: We observed one-trial differential learning in shock expectancy ratings and one-trial stimulus specificity in both shock expectancy ratings and SCR. Furthermore, the level of one-trial stimulus specificity, as indexed by the SCR, tended to correlate negatively with trait anxiety, such that higher trait-anxious individuals displayed less stimulus specificity.

The failure to observe one-trial learning in SCR may suggest that the physiological mechanism of fear conditioning is slower than the cognitive mechanism, but it could also result from procedural aspects of Experiment 1. During the test phase, the CS- (C) was scheduled to always come first, because we reasoned that the non-reinforced presentation of this stimulus would be the least surprising and therefore produce less carry-over effects to the remaining stimuli at test. However, the 20 seconds break in between conditioning and testing may have elicited an orientation response that would increase the SCR to the CS- and consequently diminish the differential CS+/CS- response (Dawson, Schell, & Filion, 2007). Experiment 2 removed this confound, in order to assess one-trial fear learning more clearly.

4. Experiment 2

In Experiment 2, we optimized the one-trial overshadowing procedure to test whether orientation responses on the first test trial indeed obscured the one-trial learning effect as indexed by SCR. First, we omitted the 20-second pause between training and test. Similar to Experiment 1, the CS- was consistently presented first during testing but this time immediately following the last training trial, limiting the orientation response. Second, fear-potentiated startle responses were measured since this psychophysiological measure is less sensitive to orientation responses (Tassinary, Cacioppo, & Vanman, 2007). Finally, we increased the number of participants to provide more power to detect fast-specific learning in our indices of fear.

4.1 Material and methods

4.1.1 Participants

Data from two nearly identical replications were merged. In total 45 psychology students (36 women) with a mean age of 19.9 (SD = 1.83) participated in Experiment 2. All participants signed informed consent and were aware that they could terminate the experiment at any moment.

4.1.2 Apparatus

The stimuli were identical to Experiment 1.

4.1.3 Measures

Questionnaires

Participants completed the state part of the State-Trait Anxiety Inventory in addition to the trait part that was completed in the first experiment (STAI-T, STAI-S; Spielberger, 1983; Dutch translation by van der Ploeg, 2000).

Fear-potentiated startle

In addition to shock expectancy ratings and skin conductance responses, fear potentiated startle was registered. Orbicularis oculi electromyographic activity (EMG) was registered through three Ag/AgCl (.25 cm diameter) electrodes filled with TECA electrolyte. After cleaning the skin with scrub cream to reduce inter-electrode resistance, three electrodes were placed on the left side of the face according to site specifications proposed by Fridlund and Cacioppo (1986). A Coulbourn isolated bioamplifier with bandpass filter (v75-04) amplified the raw signal. The recording bandwidth of the EMG signal was between 10 Hz and 20 Hz (3dB). The resulting signal was rectified online and smoothed by a Coulbourn multifunction integrator (v76-23 A) with a time constant of 50 ms. The signal was measured 100 ms before probe onset at 1000 Hz. From probe onset, the signal was digitized at 1000 Hz for 200 ms.

4.1.4 Procedure

Data of two nearly identical replications of Experiment 1 were merged. Similar to Experiment 1, stimuli B, C and X were counterbalanced as close as possible: in the final sample each stimulus combination (six in total) was used eight or nine times. In both replications the 20 seconds break between training and testing was omitted, in order to decrease the orientation response to the first stimulus presented during test (always stimulus

C). Additionally, the testing phase was extended with three additional test trials. In one of the two experiments, the presentation order of the reinforced compound AX and simple stimulus B trial was counterbalanced.

4.2 Results

4.2.1 Data reduction

Due to a programming error, seven participants had to be excluded from analysis of FPS. 15 participants did not complete the STAI-S questionnaire.

Analyses of SCR data were performed in the same way as for the first experiment. Fear potentiated startle responses were extracted with PSPHA software for analyzing psychophysiological data (De Clerck, Verschuere, De Vlieger, & Crombez, 2006). FPS amplitudes were calculated by determining the maximum EMG value between 21 ms and 175 ms after stimulus onset and subtracting the average EMG value during baseline (0 to 20 ms after stimulus onset). Resulting amplitudes were standardized into Z-scores.

4.2.2 Shock expectancy ratings

Single conditioning trial

The left panel of Figure 3A suggests that on the first and only training trial, prior to any conditioning, shock expectancies were equally strong for stimuli B, C and stimulus-compound AX. This was confirmed by a Repeated Measures (RM) ANOVA with one within-subjects factor (Stimulus, three levels) and one between-subjects factor (Replication, 2 levels), which revealed no main effect of Stimulus, $F(1.65, 69.18) < 1$, and no CS x Replication interaction, $F(1.65, 69.18) < 1$.

Test

The right panel of Figure 3A suggests higher shock expectancy ratings for B than for C and X on the first test trial. This was confirmed by a 3 (CS) x 2 (Replication) RM ANOVA which revealed a main effect of Stimulus, $F(1.42, 59.65) = 38.38, p < .001, \eta^2_{part} = .48$, and

no significant Stimulus x Replication interaction, $F(1.42, 59.65) < 1$. Planned comparisons revealed a difference in shock expectancy ratings between B and C, $F(1, 42) = 49.85, p < .001, \eta^2_{part} = .54$ (one-trial learning), and B and X, $F(1, 42) = 27.74, p < .001, \eta^2_{part} = .40$ (one-trial selective learning).

4.2.3 Skin Conductance responses

Single conditioning trial

The left panel of Figure 2B suggests equally strong SCRs for AX, B and C on the first and only training trial. This was confirmed by a 3 (Stimulus) x 2 (Replication) RM ANOVA which revealed no main effect of Stimulus, $F(1.73, 74.71) < 1$, and no Stimulus x Replication interaction, $F(1.73, 74.71) = 2.04, p = .14$.

Test

The right panel of Figure 2B suggests higher SCRs for B then for C and X on the first test trial. This was confirmed by a 3 (Stimulus) x 2 (Replication) RM ANOVA, which showed a main effect of Stimulus, $F(2, 86) = 3.49, p = .04, \eta^2_{part} = .08$, and no Stimulus x Replication interaction, $F(2, 86) < 1$. Planned comparisons revealed a difference in SCR between B and C, $F(1, 43) = 4.46, p = .04, \eta^2_{part} = .09$ (one-trial learning), and B and X, $F(1, 43) = 5.41, p = .025, \eta^2_{part} = .11$ (one-trial selective learning).

4.2.4 Fear Potentiated Startle

Single conditioning trial

The left panel of Figure 2C suggests lower FPS for AX than for B and C during training. However, a 3 (Stimulus) x 2 (Replication) RM ANOVA revealed no main effect of Stimulus, $F(1.59, 56.92) = 1.37, p = .26$, and no Stimulus x Replication interaction, $F(1.59, 56.92) < 1$.

Test

The right panel of Figure 3C suggests lower FPS for X than B, but no difference in FPS between B and C. A 3 (Stimulus) x 2 (Replication) RM ANOVA revealed no main effect of Stimulus, $F(2, 72) = 1.10, p = .34$, and no Stimulus x Replication interaction, $F(2, 72) = 1.74, p = .84$. Planned comparisons revealed no difference between B and C, $F(1, 36) < 1$, and no difference between B and X, $F(1, 36) < 1$.

Figure 3D that shows FPS for subjects with higher FPS for B than C on the test trial (one-trial learning, $N = 16$) suggests that these participants also show higher FPS for B than X (one-trial specific learning). This was confirmed by a 2 (Stimulus, B and X) x 2 (Replication) RM ANOVA which revealed a main effect of Stimulus, $F(1, 14) = 28.71, p < .001, \eta^2_{part} = .67$, and no Stimulus x Replication Interaction, $F(1, 14) < 1$.

4.2.5 Questionnaires

Similar to the first experiment, non-parametric Spearman Rho correlations were calculated between STAI-T and STAI-S on the one hand and conditioned fear to X and a selective fear index on the other hand. The selective fear index was calculated by subtracting conditioned responding to X from conditioned responding to B on the first test trial.

Correlations were tested two-sided.

STAI-T

STAI-T scores did not correlate with responding to X on the first test trial nor with the selective learning index for any of the three fear measures, all $\rho < .30, p > .14$.

STAI-S

STAI-S scores showed a marginally significant correlation with shock expectancy ratings for X on the first test trial, $\rho = .33, p = .07$. The more anxiety participants experienced during the experiment, the higher their shock expectancy for X on the first test trial. SCR and FPS for X on the first test trial did not correlate with STAI-S scores, all $\rho < .09, p > .58$.

The selective learning index did not correlate with STAI-S scores in any of the three fear measures, all $\rho < .10$, all $p > .60$.

4.3 Discussion

Using an optimized design and combining a higher number of participants, Experiment 2 unequivocally demonstrated fast-specific fear conditioning, as indexed by both shock expectancy ratings and skin conductance responses. The fear-potentiated startle results showed no evidence for fast learning overall, but those participants that did show fast learning also showed specific learning (the OTO effect). These findings replicate and extend the results from Experiment 1. In addition, individual levels of state anxiety tended to correlate positively (although not significantly) with shock expectancy ratings in the OTO test, which may support the hypothesized relevance of OTO to clinical anxiety.

5. General Discussion

The reported experiments were set up to validate the one-trial overshadowing (OTO) procedure for the study of fast-yet-specific fear learning in humans. Fear learning was *fast*, as reflected by higher shock expectancy ratings and skin conductance reactivity (SCR) to a CS+ that was paired once with an aversive electrical shock than to a CS- that was presented once without shock. At the same time, fear learning was *specific*, as reflected by lower shock expectancy, lower SCR and lower fear-potentiated startle to a stimulus that had been accompanied by another stimulus during the pairing with the shock than to a stimulus that had been paired with shock individually. This suggests that fear acquisition binds the fear reactions to the specific stimulus situation during threat. We argue that this fast-specific fear learning underlies adaptive processing of traumatic experiences, as it prevents isolated stimuli (so-called trauma reminder cues) from fully activating the trauma memory and full-blown fear reactions. Accordingly, impairments in fast-specific fear learning may lead to maladaptive

processing of traumatic experiences and ultimately contribute to chronic conditions like post-traumatic stress disorder.

Many theories can account for multiple trial overshadowing, but only a few explain one trial overshadowing as well. According to non-associative theories overshadowing results from increased ambiguity experienced during confrontations with the overshadowed stimulus. This stimulus has been paired with shock, however, always in the presence of a second stimulus. A sudden confrontation with the individual stimulus creates a situation of uncertainty, resulting in decreased conditioned responding (De Houwer, Vandorpe, & Beckers, 2005; Mitchell, De Houwer, & Lovibond, 2009). One can assume that the level of uncertainty is even larger after only a single conditioning trial. Configural associative theories explain (one trial) overshadowing in terms of a generalization decrement. These theories assume that all stimuli present are combined into a configural representation that then becomes associated with the US (see Figure 4). Future confrontations with individual stimuli will partially activate the configural representation and hence only partially activate the acquired US association, resulting in a conditioned fear reaction that is lower compared to a stimulus that would have been conditioned on its own (Pearce, 1987, 1994). The incomplete generalization mechanism from the compound to its individual stimuli is active from the first learning trial on. Consequently, configural learning theories predict that overshadowing will emerge after only a single learning trial (OTO).

Elemental associative learning theories propose an elemental processing style instead of a configural one. Elemental processing entails that all stimuli present enter the associative process individually and compete for associative strength with the US (see Figure 4). The resulting strength of the individual associations depends on the prior conditioning history of all stimuli involved and is limited by their summed associative strength (which cannot supersede the total associative strength supported by the US). Unlike a stimulus that was

conditioned individually, the gain in associative strength for a stimulus that was conditioned in compound is not only limited by its own associative strength from its previous conditioning history, but also by the associative strength of the other stimuli in the compound. Future confrontations with individual stimuli will activate only the partial association acquired by that stimulus, and hence elicit a conditioned fear reaction that is lower compared to a stimulus that would have been conditioned on its own (not limited by summed associative strength during the associative process). However, in the case of a *single* conditioning trial, all stimuli are associatively neutral when they enter the associative process and so acquisition of individual associative strength is not yet limited by summed associative strength (which is zero at that moment). Thus, on that very first compound conditioning trial, each individual stimulus in the stimulus-combination gains the same amount of associative strength as a stimulus conditioned individually. Hence, future confrontations with the individual stimuli will elicit a conditioned fear reaction that is equal to a stimulus that would have been conditioned on its own and no overshadowing effect emerges.

Research into the neurobiology of configural learning has emphasized the role of two antagonistic processes in the hippocampus: pattern completion and pattern separation. Pattern completion is computed in the CA3 region of the hippocampus and promotes full activation of a stored configuration (“pattern”) from partial input. It supposedly serves as a neurobiological basis of generalization, by processing a test situation similarly to a learning situation despite differences in their input patterns. Dysregulation of pattern completion may also lead to overgeneralization, as in the case of PTSD where a single cue can reactivate the trauma memory in a highly vivid way that leads to a full-blown fear reaction (Kheirbek, Klemenhagen, Sahay, & René-Hen, 2012; Leutgeb, Leutgeb, Moser, & Moser, 2007; Lissek, 2012). Pattern separation is computed in the dentate gyrus of the hippocampus and promotes separation of similar/overlapping input patterns into distinct memory representations. It

supposedly serves as a neurobiological basis of discrimination, by processing a test situation differently from a learning situation. Creating separate memory representations reduces the probability of pattern overlap and impedes the initiation of pattern completion (Kheirbek et al., 2012; Leutgeb, Leutgeb, Moser, & Moser, 2007; Lissek, 2012). Impaired pattern separation may give way to erroneous pattern completion and produce overgeneralization. In the case of PTSD, fear overgeneralization may result from impaired pattern separation during confrontations with a trauma reminder in a safe situation, giving way to erroneous pattern completion and fear overgeneralization (Kheirbek et al., 2012; Lissek, 2012). In conclusion, where associative learning theory stresses the importance of configural versus non-configural (elemental) processing during the learning itself, pattern completion/separation theory adds a level of analysis by pointing out how memory processes determine the extent to which stored configurations can lead to generalization/discrimination during testing. In the case of OTO, this means that increased levels of fear could still follow configural processing, if the pattern completion/separation process is dysregulated at test.

There are a number of limitations to this study. First, the different fear measures dissociated somewhat with regard to fast-specific learning: fast simple discriminatory learning was observed in shock expectancy ratings and SCR, but not FPS. Moreover, FPS only showed fast stimulus selective learning in participants that also exhibited fast simple discriminatory learning. These results suggest that FPS is less sensitive measure for one-trial learning effects than shock expectancy ratings or skin conductance responses. Another possibility is that different measures reflect different response components of the fear construct, each with a different learning rate. Previous research showed a close association between shock expectancy data and SCR, suggesting that each reflects declarative knowledge or contingency learning (Soeter & Kindt, 2010). These measures often dissociate from FPS, which is considered to reflect the emotional fear response (Soeter & Kindt, 2010). These results

suggest that contingency learning occurs faster than emotional learning, resulting in a learning effect after only a single learning-trial in measures of stimulus contingencies (SCR and shock expectancy ratings) but not the emotional fear response (FPS). Second, the correlations with anxiety questionnaires were marginally significant and not consistent between the two experiments. Experiment 1 complements earlier findings from Boddez et al. (2012), who observed a positive correlation between trait anxiety and shock expectancy for part of a reinforced stimulus configuration. However, in Experiment 2 stimulus-specificity tended to correlate (although not significantly) with state anxiety and not trait anxiety. State anxiety was not assessed in Experiment 1 or the study of Boddez et al. (2012). Possibly, assessing state anxiety influences trait anxiety scores, unless both aspects are assessed separately. These findings raise the question whether state anxiety and not trait anxiety is the crucial vulnerability factor for impaired stimulus specific learning. Third, it is unclear to what extent the different stimuli in our experiments were associatively neutral during the conditioning trial. Pre-experimental instructions informed participants that they would receive electrical shocks during the experiment and that some stimuli would be followed by the electrical shock. This may have installed shock expectancies for all stimuli previous to the single conditioning trial (Mechias, Etkin, & Kalisch, 2010). If this is the case, the observed overshadowing is in fact a multiple trial learning effect that can be explained by elemental conditioning theories after all (from an associative learning perspective). Ethical considerations preclude the possibility of administering shocks in the absence of any instructions. However, future studies could add a phase prior to the conditioning trial where all stimuli are presented a number of times without shock in order to reduce any instruction-based expectancy of shock. Fourth, the current experimental design does not allow disentangling an associative explanation of OTO in terms of elemental/configural processing from non-associative explanations in terms of uncertainty/ambiguity (based on probabilistic

and propositional models of learning, De Houwer, Vandorpe, & Beckers, 2005; Mitchell, De Houwer, & Lovibond, 2009). Future studies could measure individual differences in uncertainty/ambiguity tolerance (e.g., the intolerance of uncertainty scale, Freeston et al., 1994) as well as in configural processing abilities (see Gilbertson et al. 2007), and compare their correlational strengths with OTO. Also, manipulations that are known to bias participants towards elemental processing should affect OTO *only* if elemental-configural processing is the underlying mechanism of OTO.

6. Conclusion

A one-trial overshadowing procedure was validated to assess fast-specific fear learning in humans. Results revealed that a single conditioning trial produced (1) increased shock expectancy and SCR, compared to a non-conditioned control stimulus (fast learning), and (2) decreased shock expectancy, SCR and FPS to a stimulus that was part of a conditioned stimulus compound (fast-specific learning). The level of fast-specific learning correlated marginally significantly with (trait and state) anxiety scores. We conclude that OTO is a promising tool to study fast-specific learning in anxiety patients and examine mechanisms involved in the development of fear overgeneralization.

Acknowledgements

This research was supported by research grant G.0942.12N of the Research Foundation – Flanders (FWO), KU Leuven program funding grant PF/10/005 and FP7 Marie Curie International Outgoing Fellowship 627743. We would like to thank the participants of this study.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
- Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological psychology*, 92(1), 90-96.
- Boddez, Y., Vervliet, B., Baeyens, F., Lauwers, S., Hermans, D., & Beckers, T. (2012). Expectancy bias in a selective conditioning procedure: trait anxiety increases the threat value of a blocked stimulus. *Journal of behavior therapy and experimental psychiatry*, 43(2), 832-837.
- Cole, R. P., Oberling, P., & Miller, R. R. (1999). Recovery from one-trial overshadowing. *Psychonomic bulletin & review*, 6(3), 424-431.
- Davey, G. C. (1992). Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behaviour Research and Therapy*, 14(1), 29-66.
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds), *Handbook of Psychophysiology*, 3rd edition (pp. 159-181), UK: Cambridge University Press.
- De Clercq, A., Verschuere, B., De Vlieger, P., & Crombez, G. (2006). Psychophysiological Analysis (PSPHA): A modular script-based program for analyzing psychophysiological data. *Behavior research methods*, 38(3), 504-510.
- De Houwer, J., Vandrope, S., & Beckers, T. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior*, 33(2), 239-249.

568 Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear
569 generalization in humans: systematic review and implications for anxiety disorder
570 research. *Behavior therapy*, 46(5), 561-582.

571 Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder.
572 *Behaviour research and therapy*, 38(4), 319-345.

573 Ehlers, A., Hackmann, A., Steil, R., Clohessy, S., Wenninger, K., & Winter, H. (2002). The
574 nature of intrusive memories after trauma: The warning signal hypothesis. *Behaviour*
575 *research and therapy*, 40(9), 995-1002.

576 Fanselow, M. S. (1990). Factors governing one-trial contextual conditioning. *Animal*
577 *Learning & Behavior*, 18(3), 264-270.

578 Freeston, M. H., Rheaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do
579 people worry? *Personality and Individual Differences*, 17, 791-802.

580 Gilbertson, M. W., et al. (2007). Configural cue performance in identical twins discordant for
581 posttraumatic stress disorder: theoretical implications for the role of hippocampal
582 function. *Biological Psychiatry*, 62, 513-520.

583 Hull, C. L. (1943). *Principles of Behaviour*. New-York: Appleton-Century-Crofts.

584 Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research.
585 *Psychophysiology*, 23(5), 567-589.

586 James, J. H., & Wanger, A. R. (1980). One-trial overshadowing: Evidence of distributive
587 processing. *Journal of Experimental Psychology: Animal Behavior Processes*, 6(2),
588 188.

589 Jarvik, M. E., & Essman, W. B. (1960). A simple one-trial learning situation for mice.
590 *Psychological Reports*, 6(2), 290-290.

591 Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin*
592 *of the Psychonomic Society*, 18(6), 318-320.

593 Kaye, H., Gambini, B., & Mackintosh, N. J. (1988). A dissociation between one-trial
 594 overshadowing and the effect of a distractor on habituation. *The Quarterly Journal of*
 595 *Experimental Psychology*, 40(1), 31-47.

596 Kheirbek, M. A., Klemenhagen, K. C., Sahay, A., & Hen, R. (2012). Neurogenesis and
 597 generalization: a new approach to stratify and treat anxiety disorders. *Nature*
 598 *neuroscience*, 15(12), 1613-1620.

599 Lanzetta, J. T., & Orr, S. P. (1980). Influence of facial expressions on the classical
 600 conditioning of fear. *Journal of personality and social psychology*, 39(6), 1081.

601 Lanzetta, J. T., & Orr, S. P. (1981). Stimulus properties of facial expressions and their
 602 influence on the classical conditioning of fear. *Motivation and Emotion*, 5(3), 225-
 603 234.

604 Leutgeb, J. K., Leutgeb, S., Moser, M. B., & Moser, E. I. (2007). Pattern separation in the
 605 dentate gyrus and CA3 of the hippocampus. *science*, 315(5814), 961-966.

606 Lissek, S. (2012). Toward an account of clinical anxiety predicated on basic, neurally mapped
 607 mechanisms of Pavlovian fear-learning: The case for conditioned overgeneralization.
 608 *Depression and anxiety*, 29(4), 257-263.

609 Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon,
 610 C. (2008). Generalization of conditioned fear-potentiated startle in humans:
 611 experimental validation and clinical relevance. *Behaviour research and therapy*, 46(5),
 612 678-687.

613 Lissek, S., Pine, D. S., & Grillon, C. (2006). The strong situation: A potential impediment to
 614 studying the psychobiology and pharmacology of anxiety disorders. *Biological*
 615 *psychology*, 72(3), 265-270.

616 Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., &
 617 Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-
 618 analysis. *Behaviour research and therapy*, 43(11), 1391-1424.

619 Mackintosh, N. J. (1971). An analysis of overshadowing and blocking. *The Quarterly Journal*
 620 *of Experimental Psychology*, 23(1), 118-125.

621 Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal learning &*
 622 *behavior*, 4(2), 186-192.

623 Mackintosh, N. J., & Reese, B. (1979). One-trial overshadowing. *The Quarterly Journal of*
 624 *Experimental Psychology*, 31(3), 519-526.

625 McNally, G. P., & Westbrook, R. F. (2003). Anterograde amnesia for Pavlovian fear
 626 conditioning and the role of one trial overshadowing: Effects of preconditioning
 627 exposures to morphine in rat. *Journal of Experimental Psychology: Animal Behavior*
 628 *Processes*, 29(3), 222.

629 Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies:
 630 implications for conscious appraisal of threat. *Neuroimage*, 49(2), 1760-1768.

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

633 Öhman, A., Eriksson, A., & Olofsson, C. (1975). One-trial learning and superior resistance to
 634 extinction of autonomic responses conditioned to potentially phobic stimuli. *Journal*
 635 *of Comparative and Physiological Psychology*, 88(2), 619.

636 Pavlov, I. P. (1927). Conditioned reflexes. An Investigation of the physiological activity of
 637 the cerebral cortex.

638 Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning.
 639 *Psychological review*, 94(1), 61.

640 Pearce, J. M. (1994). Similarity and discrimination: a selective review and a connectionist
641 model. *Psychological review*, 101(4), 587.

642 Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning. Variations in
643 the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F.
644 Prokasy (Eds.), *Classical Conditioning II: Current research and theory*. New York:
645 Appleton Century Crafts.

646 Rudy, J. W., & O'Reilly, R. C. (1999). Contextual fear conditioning, conjunctive
647 representations, pattern completion, and the hippocampus. *Behavioral neuroscience*,
648 113(5), 867.

649 Rudy, J. W., & O'Reilly, R. C. (2001). Conjunctive representations, the hippocampus, and
650 contextual fear conditioning. *Cognitive, Affective, & Behavioral Neuroscience*, 1(1),
651 66-82.

652 Soeter, M., and Kindt, M. (2010). Dissociating response systems: erasing fear from memory.
653 *Neurobiology of Learning and Memory*, 94, 30-41.

654 Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory STAI (form Y)("
655 self-evaluation questionnaire").

656 Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0:
657 A free software package for implementing psychological and psychophysiological
658 experiments. *Experimental Psychology*, 57, 36-45.

659 Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The Skeletomotor System:
660 Surface. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds), *Handbook of*
661 *Psychophysiology*, 3rd edition (pp. 159-181), UK: Cambridge University Press.

662 van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1981). *Handleiding bij de zelf-*
663 *beoordelings vragenlijst: een Nederlandstalige bewerking van Spielberger state-trait*
664 *anxiety inventory STAI-DY. Addendum 1981*. Swets & Zeitlinger.

Zelikowsky, M., & Fanselow, M. S. (2010). Opioid regulation of Pavlovian overshadowing in fear conditioning. *Behavioral neuroscience*, 124(4), 510.

Figure Captions

Fig. 1. The experiment consists of two phases. During training (see panel 1A) a compound of two stimuli (AX) is presented once followed by an electrical shock. Two additional stimuli are presented individually; one of them is followed by shock (B), the other one is not (C). During test (see panel 1B) each stimulus from the training phase is presented individually, without shock. Shock is always delivered 7.5 s following stimulus onset. To measure shock expectancy an eleven-point scale appears below every stimulus (and stays on screen as long as the stimulus or until a rating is given). To measure startle responses in Experiment 2, a startle probe is presented on every trial 7 s following stimulus onset. The stimulus disappears from the screen after 8 seconds. A novel stimulus appeared on screen 12-16 seconds later.

Fig. 2. Mean shock expectancy ratings (2A), skin conductance responses (SCR, 2B) for training and the first test trial are shown. Shock expectancy ratings at test are higher for B (black bar) than for C (white bar, one trial learning) and X (grey bar, one trial selective learning). SCR at test are higher for B than for X, but not C. FPS at test does not differ between B, C and X. Figure 2D shows FPS for training and the first test trial for participants that showed higher FPS for B than C. These participants showed higher FPS for B (black bar) than X (grey bar, one trial selective learning). Error bars represent standard error of means.

Fig. 3 Mean shock expectancy ratings (3A), skin conductance responses (SCR, 3B) and Fear Potentiated Startle (FPS, 3C) for training and the first test trial are shown. Shock expectancy ratings at test are higher for B (black bar) than for C (white bar, one trial learning) and X (grey bar, one trial selective learning). SCR at test are higher for B than for X, but not C. FPS at test does not differ between B, C and X. Figure 3D shows FPS for training and the first test trial for participants that showed higher FPS for B than C (N = 16). These participants showed

higher FPS for B (black bar) than X (grey bar, one trial selective learning). Error bars represent standard error of means.

Fig 4. For a single conditioning trial with a stimulus configuration, two memory structures are possible according to associative learning theories. When the configuration is processed elementally; each stimulus is associated with the outcome individually. The associative strength each stimulus gains depends on the previous learning history of all the stimuli in the configuration. On a single learning trial all stimuli are associatively neutral when they enter the associative process. Learning is not limited by the presence of other stimuli in the configuration. Hence, during test, the individual stimulus will elicit a conditioned fear reaction equal to a stimulus that would have been conditioned on its own. The configuration can also be processed configurally. All stimuli are combined into a configural representation that becomes associated with the US. During test, the individual stimulus will only partially activate the acquired US association, resulting in a conditioned fear reaction that is lower compared to a stimulus that would have been conditioned on its own (Pearce, 1987, 1994).

740
741
742
743
744
745

746
747
748

Table 1

Overview of the experimental phases.

Training	Test
AX + (1)	X - (3/6)
B+ (1)	B - (3/6)
C- (1)	C - (3/6)

Note: A, X, B and Y are geometrical shapes (X, B and Y counterbalanced); '+' refers to the administration of the shock US; '-' refers to the absence of shock. The first number between parentheses represents the number of trials for Experiment 1, the second number represents the number of trials for Experiment 2.

Figure(s)
A Training

B Test

Figure(s)

Figure(s)

A

B

C

D

Figure(s)

PREVIOUS LEARNING = 0

Single Conditioning Trial

Elemental Processing

Configural Processing

Test

FEAR

FEAR