Pavlovian Occasion Setting in Human Fear and Appetitive Conditioning: Effects of Trait Anxiety and Trait Depression

Tomislav D. Zbozinek, Ph.D.[a]

Toby Wise, Ph.D. [a]

Omar Perez, Ph.D. [a,b]

Song Qi, Ph.D. [c]

Michael S Fanselow, Ph.D.[d]

Dean Mobbs, Ph.D. [a]

a. California Institute of Technology, Humanities and Social Sciences, 1200 E. California Blvd., MC 228-77, Pasadena, CA 91125, USA

b. University of Santiago of Chile (USACH), Faculty of Business and Economics, Santiago, Chile.

c. National Institute of Mental Health, 6001 Executive Boulevard, Room 6200, MSC 9663 Bethesda, MD 20892

d. University of California, Los Angeles, Department of Psychology, 502 Portola Plaza, Los Angeles, CA 90025

Correspondence to:
Tomislav D. Zbozinek, Ph.D.
Division of Humanities and Social Sciences
California Institute of Technology
1200 E. California Blvd., MC 228-77
Pasadena, CA 91125, USA
Email: zbozinek@caltech.edu

Abstract

Contexts and discrete stimuli often influence the association between a stimulus and outcome. This phenomenon, called occasion setting, is central to modulation-based Pavlovian learning. We conducted two experiments with humans in fear and appetitive conditioning paradigms, training stimuli in differential conditioning, feature-positive discriminations, and feature-negative discriminations. We also investigated the effects of trait anxiety and trait depression on these forms of learning. Results from both experiments showed that participants were able to successfully learn which stimuli predicted the electric shock and monetary reward outcomes. Additionally, as hypothesized, the stimuli trained as occasion setters had little-to-no effect on simple reinforced or non-reinforced stimuli, suggesting the former were indeed occasion setters. Lastly, in fear conditioning, trait anxiety was associated with increases in fear of occasion setter/conditional stimulus compounds; in appetitive conditioning, trait depression was associated with lower expectations of monetary reward for the trained negative occasion setting compound and transfer of the negative occasion setter to the simple reinforced stimulus. These results suggest that clinically anxious individuals may have enhanced fear of occasion setting compounds, and clinically depressed individuals may expect less reward with compounds involving the negative occasion setter.

Keywords: fear conditioning; reward conditioning; occasion setting; anxiety; depression; Pavlovian conditioning

Evolution works to organize biological systems that are important to survival, including how to predict and respond to ecological threats (Mobbs et al., 2020) and how to optimize reward-related behavior (O'Doherty et al., 2017). A major way to predict reward/danger is by learning Pavlovian stimulus-outcome associations, where a conditional stimulus (CS) may signal the presence (CS+) or absence (CS-) of a biologically relevant outcome (i.e., the unconditional stimulus; US). In fear conditioning, the US is an aversive stimulus (e.g., electric shock), which makes the CS+ a danger signal and the CS- a safety signal. In appetitive conditioning, the US is a rewarding stimulus (e.g., food, money), which makes the CS+ a reward signal and the CS- a signal for non-reward. While fundamental to learning, these simple CS+/US and CS-/No US associations fail to capture some of the complexities encountered by organisms in naturalistic settings, where the relationships between stimuli can be context-dependent or influenced by other stimuli.

Pavlovian occasion setting is a modulatory form of learning in which the CS's ability to signal the US depends on the presence or absence of the occasion setter (Bonardi et al., 2017; Fraser & Holland, 2019). In simple Pavlovian learning, the CS usually signals both whether and when the US will occur. In occasion setting, the "whether" and "when" are divided: the occasion setter signals whether the CS will result in the US, and the CS signals when the US will occur (Fraser & Holland, 2019). Occasion setters can be contexts (e.g., time, place) or discrete stimuli (e.g., tone, light) (Fraser & Holland, 2019; Trask et al., 2017), and occasion setters often onset prior to the CS since simultaneous presentation with the CS often leads to direct associative learning rather than occasion setting (Fraser & Holland, 2019; Holland, 1986a). There are two forms of occasion setting: positive occasion setting occurs when the CS predicts the US *only if* the positive occasion setter was recently presented, and negative occasion setting occurs when the CS predicts the US *unless* the negative occasion setter was recently presented.

It is important to experimentally determine whether a given stimulus is a CS or an occasion setter, which can be done by assessing the degree to which the putative occasion setter affects a CS it was not trained with (i.e., how well it transfers to that CS) (Bonardi et al., 2017; Fraser & Holland, 2019; Trask et al., 2017). Specifically, an occasion setter will only transfer to a CS that has undergone similar training with an occasion setter (Baeyens et al., 2001, 2004; Holland, 1986b, 1989a, 1989b, 1991b, 1991a), so it will have little-to-no effect on responding to a consistently treated CS+ or CS- (Baeyens et al., 2004; Holland, 1986b, 1989b; Holland et al., 1999; Holland & Lamarre, 1984; Lamarre & Holland, 1985, 1987). We conduct this type of transfer test in the present report.

Additionally, occasion setting has received very little attention in human research (Baeyens et al., 2001, 2004; Balea et al., 2020; De Houwer et al., 2005; Declercq & De Houwer, 2008; Dibbets et al., 2002; Franssen et al., 2017; Ruprecht et al., 2014; van Vooren et al., 2012). Most of the occasion setting experiments in humans are done with aversive conditioning (Baeyens et al., 2001, 2004; Balea et al., 2020; Dibbets et al., 2002; van Vooren et al., 2012), and even the "aversive" stimuli are usually fairly mild (e.g., losing points or playing a flashing screen with sound pattern, as opposed to electric shocks; De Houwer et al., 2005). There are very few appetitive occasion setting experiments in humans; the existing ones utilize gaining or losing points (Dibbets et al., 2002) or obtaining treasure chests (Ruprecht et al., 2014) based on an instrumental response. We are unaware of an appetitive occasion setting study in humans using purely Pavlovian procedures; indeed, most of the aversive human occasion setting studies are instrumental, as well (Baeyens et al., 2001, 2004; Balea et al., 2020; De Houwer et al., 2005; Franssen et al., 2017; van Vooren et al., 2012). Thus, there is a dearth of human research on a) appetitive occasion setting (Pavlovian and instrumental), b) Pavlovian aversive occasion setting, and c) aversive occasion

setting using traditional aversive stimuli (e.g., electric shock). Our present report adds to the human occasion setting literature by conducting aversive Pavlovian occasion setting with electric shocks (Study 1) and appetitive Pavlovian conditioning using monetary rewards (Study 2).

Furthermore, because real-world Pavlovian associations are likely more complicated than those experienced in simple Pavlovian laboratory experiments, occasion setting may be important for understanding general associative learning in humans and improving our understanding and treatment of mental health disorders (e.g., anxiety, depression). Using an anxiety example, the degree to which a client with fear of public speaking expects the audience (CS) to reject them (US) often depends on situational factors, such as the physical setting (at a party or a conference) or whether they give their speech immediately after a particularly charismatic and engaging speaker. Similarly, using a depression example, the degree to which a client with anhedonia expects listening to music (CS) to be enjoyable (US) may depend on situational factors, such as the physical setting (at a concert or at home) or whether they are listening to music after friends and family recently visited or while home alone all day. Many of these situational factors are likely occasion setters that modulate whether the CS will result in the US.

Occasion setting likely has treatment implications for anxiety and depression. With anxiety disorders, it is critical to distinguish occasion setters from CSs in exposure therapy because conducting extinction/exposure to the positive occasion setter alone will not affect its ability to signal whether the CS predicts the US (e.g., Rescorla, 1986). Using the anxiety example above, conducting exposures to just watching a charismatic and engaging speaker would not reduce that speaker's ability to increase the client's fear of giving a speech. However, conducting exposures to giving a speech after the charismatic speaker would extinguish both fear of the client's speech and the charismatic speaker's ability to signal that the client's speech will result in rejection

(Franssen et al., 2017; Miller & Oberling, 1998; Rescorla, 1986; van Vooren et al., 2012). While the inhibitory retrieval model of exposure therapy includes principles of negative occasion setting, such as extinction/exposure in multiple contexts to generalize negative occasion setting learning (Craske et al., 2014), exposure therapy has not yet formally incorporated positive occasion setting principles. This is a subtle but potentially impactful approach that could enhance treatment outcome and reduce return of fear and relapse (Craske & Mystkowski, 2006; Scholten et al., 2016; van Dis et al., 2020) since an extinguished CS paired with an unextinguished positive occasion setter will produce fear (assuming the CS has been trained with a positive occasion setter previously). Additionally, a significant feature of anxiety disorders is elevated fear of both safe and dangerous stimuli (Craske et al., 2012; Duits et al., 2015; Dymond et al., 2015; Jovanovic et al., 2012; Lissek, 2012; Lissek et al., 2005, 2008, 2010, 2014). Meta-analyses have found that individuals with anxiety disorders demonstrated greater fear of both the CS+ (acquisition and extinction) and CS- (acquisition) compared to healthy controls (Duits et al., 2015; Lissek et al., 2005). Other studies have found that trait-anxious individuals display poorer inhibitory learning (Grillon & Ameli, 2001; Kindt & Soeter, 2014; Laing et al., 2021; Staples-Bradley et al., 2018), increased excitatory threat of a blocked stimulus (Boddez et al., 2012), and greater fear of trained stimuli (Chan & Lovibond, 1996) and generalization stimuli (Wong & Lovibond, 2018) when the rule for the CS/US association is uncertain. However, we are unaware of any studies that have investigated the effects of anxiety on fear during positive and negative occasion setting in fear conditioning.

Moreover, with depression, a lack of understanding of the conditions for specific situations or behaviors to result in rewarding outcomes may mitigate the effects of treatment (e.g., behavioral activation). Using the depression example above, simply listening to music might not be a reliable

way to experience enjoyment, but listening to music in specific contexts or under certain conditions may be more rewarding (e.g., after friends and family visited). In the treatment of depression, there is no explicit focus on occasion setting, which is probably a result of little-to-no research examining the potential association between depression and occasion setting. However, there is a plausible link between depression and appetitive occasion setting because depression has been associated with deficits in reward processing (Admon & Pizzagalli, 2015) and reward learning (Kumar et al., 2008, 2018; Vrieze et al., 2013). Whether this extends to occasion setting has not been tested yet. By investigating occasion setting with anxiety and depression, we can better understand associative learning mechanisms in humans and potentially improve our understanding and treatment of anxiety and depressive disorders.

In the present report, we investigate whether trait anxiety and trait depression are associated with aversive and appetitive differential conditioning and occasion setting. We conducted two conditioning experiments: fear conditioning (electric shock US) and appetitive conditioning (monetary US). These experiments are matched in many regards. Participants engaged in conditioning with six stimuli: differential conditioning (a+, b-), feature-positive training (C→d+, d-), and feature-negative training (E→f-, f+) (where "feature" indicates a putative occasion setter, "target" indicates the CS that may coincide with the US, and "→" indicates an inter-stimulus trace interval). The following are our hypotheses: participants will demonstrate greater responding to a+ than b- (Hypothesis 1a), C→d+ than d- (Hypothesis 1b), and f+ than E→f- (Hypothesis 1c). Hypothesis 2: as a test of whether occasion setting was learned (Baeyens et al., 2001, 2004; Holland, 1986b, 1989a, 1989b, 1989c, 1991b, 1991a), participants will have little-to-no transfer of the putative occasion setters (i.e., stimuli C and E) to the differential conditioning stimuli that were trained alone (i.e., stimuli a+ and b-). In Experiment 1, this will be demonstrated by more

similar responding between 1) E→a and a+ than E→a and E→f-, and 2) C→b and b- than C→b and C→d+. Experiment 2 had more transfer test stimuli, so we hypothesized this would be demonstrated by C→b and Cb < ab < E→a and Ea, where two stimuli listed together (e.g., Cb, ab, Ea) indicates simultaneous presentation. Hypothesis 3: In Experiment 1, participants with high trait anxiety (compared to low trait anxiety) will have greater fear of safe stimuli (i.e., stimuli that are not paired with shock: b-, d-, and E→f-), as this is a risk factor specific to anxiety disorder onset (Craske et al., 2012). However, it is noteworthy that clinically anxious individuals show greater fear of both safe and dangerous stimuli per meta-analyses (Duits et al., 2015; Lissek et al., 2005), suggesting trait anxiety may be related to greater fear of both. We will also investigate in exploratory analyses whether individuals high in trait anxiety (compared to low trait anxiety) have greater fear of CSs that were trained with putative occasion setters due to their more ambiguous/mixed association with the US (relative to a simple CS+ or CS-). We will also examine the effects of trait anxiety in appetitive conditioning (to assess specificity of aversive vs appetitive conditioning), as well as the effects of trait depression in aversive and appetitive conditioning. Depression is associated with deficits in reward learning (Kumar et al., 2008, 2018; Vrieze et al., 2013), so one possibility is that we will observe effects of trait anxiety in fear conditioning and trait depression in appetitive conditioning.

Experiment 1

Experiment 1 Methods

Participants

Participants (N = 80) were students/staff from the California Institute of Technology or members of the nearby community who were paid $50 for their participation. Participants were 55% female, 40% male, 1.3% female-to-male transgender, 1.3% agender, and 1.3% other; mean age 28.99 years (SD = 9.19); and 3.8% Black or African-American, 31.3% Asian or Asian-American, 15.0% Hispanic/Latinx, 37.5% White, and 11.4% Multiracial or Other. Participants were recruited via Caltech's SONA Systems, email advertisements at Caltech, or word of mouth. In order to ensure variability of trait anxiety in our sample, participants were recruited based on high/low trait anxiety as measured by the Overall Anxiety Severity and Impairment Scale (OASIS; Campbell-Sills et al., 2009; Norman et al., 2006) and Depression, Anxiety, and Stress Scale – 21-Item Version Trait Anxiety Subscale (DASS-21; Antony & Bieling, 1998; Lovibond & Lovibond, 1995). An OASIS of ≥ 8 is indicative of an individual likely having a clinical anxiety disorder (Campbell-Sills et al., 2009), and a DASS-21 anxiety score of ≥ 10 is indicative of moderate or greater anxiety. "Low Anxiety" participants were admitted into the study if they had <8 on OASIS and <10 on DASS-21 anxiety (n = 40), whereas "High Anxiety" participants were admitted into the study if they had greater than or equal to one of those scores (n = 40). This study was approved by the California Institute of Technology Institutional Review Board, and all participants provided informed consent prior to commencing the study.

The sample size needed to observe effects of trait anxiety on fear to reinforced vs non-reinforced stimuli was obtained by a power analysis prior to the study (see pre-registration for details: https://osf.io/df7jq). From this analysis, we estimated that 78 participants would be needed

to observe our effects with a power of .95 at p = .05. Due to technical difficulties that resulted in missing data on some trials for some participants, we admitted 80 participants into the study to achieve at least 78 observations per trial.

Design

Participants were presented with two within-subjects conditions: Trial (1, 2, etc.) and Stimulus (during Training/Reminder: a+, b-, C→d+, d-, E→f-, f+; during Transfer Test: C→b, E→a). Our design closely matched previous occasion setting work in training and transfer tests (e.g., Holland et al., 1999) while adapting it for humans. For Stimulus, a+/b- were differential conditioning; C→d+/d- were feature-positive training; and E→f-, f+ were feature-negative training. Trait Anxiety and Depression were measured by self-report questionnaires at the beginning of the experiment prior to fear conditioning. We used three measures of fear during conditioning: US expectancy, self-report fear, and skin conductance response.

Materials and Apparatus

The Pavlovian conditioning procedure was programmed using PsychoPy 3.2.3 (Peirce et al., 2019). See Figure 1 for trial design. CSs were 8 sec images on a computer screen. Differential conditioning CSs (i.e., a+, b-) were a green star and blue triangle (counterbalanced), CSs from feature-positive and feature-negative training (i.e., d, f) were a black square and white circle, respectively, and the putative occasion setters were 8-sec 66-71 dB auditory stimuli (counterbalanced): a trumpet sound (130 Hz, C3) and violin sound (784 Hz, G5). We chose different modalities between the putative occasion setters and CSs because using different stimulus modalities (e.g., auditory, visual) facilitates occasion setting learning, whereas using the same

modality facilitates direct associations with the US (Holland, 1989a). The US was electric shock and was delivered using STMISOC with two LEAD110A (BIOPAC, Inc.) and two Telectrode T716 Ag/AgCl electrodes. The shock consisted of two pulses .03 sec apart delivered to the underside of the wrist approximately 1-2 inches below the palm 7 sec after CS onset during reinforced trials. Intertrial intervals (ITIs) varied between 20, 25, and 30 seconds (average of 25) and consisted of a gray screen with a black fixation cross in the middle. During compound trials (e.g., C→d+, E→f-), the putative occasion setter auditory stimulus was delivered through headphones while the computer screen continued to look identical to an ITI.

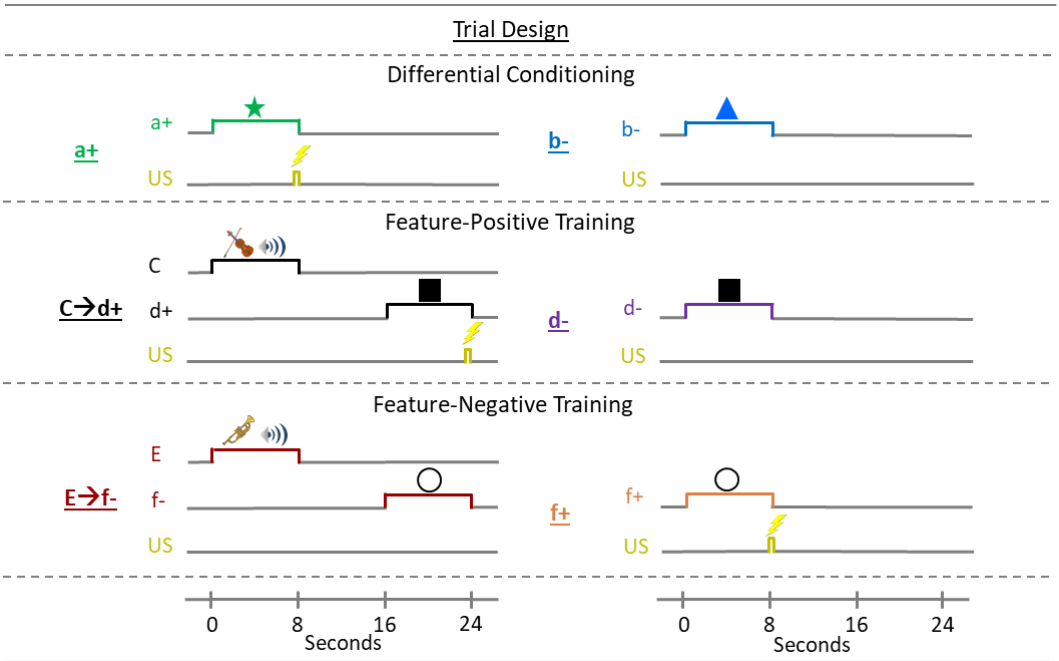

Figure 1. Trial Design. Conditions/stimuli are within-subjects. Conditional stimuli (CSs; a, b, d, f). Putative occasion setters (C, E). Differential Conditioning (a+, b-), Feature-Positive Training (C→d+, d-), and Feature-Negative Training (E→f-, f+). Unconditional stimulus (US): electric shock. On reinforced trials, US occurred 7 sec after CS onset. All CSs, occasion setters, and trace periods (i.e., time gap between occasion setter offset and CS onset) were 8 sec. Importantly, auditory stimuli were not shown on screen but are presented with images in the figure to represent the sound (i.e., violin with speaker icon, trumpet with speaker icon). All stimuli are color-coded throughout the paper: a+, b-, **C→d+**, d-, E→f-, and f+. Intertrial intervals (ITIs) were 20-30 sec. Auditory stimuli and trace periods visually looked identical to ITIs but were 8 sec.

US expectancy.

A BIOPAC MP150 hardware unit and AcqKnowledge version 5.0 software (BIOPAC Systems, Inc.) were used to acquire US expectancy and skin conductance data. Participants used a sliding scale (BIOPAC model TSD115) continuously throughout the experiment to rate US expectancy in real time during all stimuli. The instructions were, "Please rate how certain you are that you will receive electric shock in the next few moments." The values ranged from 0 = "Certain shock will not occur", 4.5 = "Uncertain," and 9 = "Certain shock will occur." US expectancy was calculated as the mean rating 6.5-7 seconds after stimulus onset.

Skin conductance response (SCR).

SCR was recorded as a measure of arousal from two EL507 11mm diameter Ag/AgCl electrodes placed on the distal phalanx of the index and middle fingers of the non-dominant hand (e.g., (Bradley et al., 1990)). Using a GSR100C amplifier and two LEAD110A, SCR data was sampled at a rate of 31.25 Hz and filtered using a low pass filter with a frequency cutoff fixed at 1Hz. SCR was calculated as a difference score between the maximum skin conductance value 1 to 6 seconds after stimulus onset minus the mean skin conductance value of the 2 seconds prior to stimulus onset. This difference score was then square-root-transformed to normalize the data, as well as range-corrected by dividing the difference score by the largest observed skin conductance response. SCRs less than zero were coded as zero.

Self-Report Fear.

To rate fear of the CSs when presented alone (i.e., stimuli A, B, D, F), participants rated "How fearful do you feel when you see this image BY ITSELF (i.e., NOT when it appears a few seconds after a sound)?" using a 1-9 scale, where 1 = "Not at all fearful", 5 = "Moderately fearful," and 9 = "Very fearful." To rate fear of the CSs that occurred after the putative occasion setters (i.e., Training and Reminder: C→d+, E→f-; Transfer Test: E→a, C→b), participate rated "How fearful do you feel when you see this IMAGE a few seconds AFTER hearing this sound?" using the same 1-9 scale. Images and sounds were presented during the rating and terminated upon completion of the fear rating; sounds otherwise terminated after 8 seconds. Self-report fear was measured before each training phase (i.e., differential conditioning, feature-positive training, feature-negative training) and after every two blocks of trials during the training phase (one block = one reinforced stimulus presentation and two non-reinforced stimulus presentations; e.g., one C→d+, two d-) for all stimuli presented in each phase. Self-report fear was also measured after Habituation, after Reminder, and after Transfer Test for all stimuli presented during those phases.

Trait Anxiety (i.e., "Anxiety").

Our primary measure of anxiety was a composite score, which was used dimensionally (not categorically) to predict fear during fear conditioning. The composite score includes subscales from several questionnaires that have previously been shown to measure the construct of anxiety. These measures include the State-Trait Anxiety Inventory – Trait Version Anxiety subscale (STAI; Bieling et al., 1998; Spielberger & Gorsuch, 1983); the Depression, Anxiety, and Stress Scale 21-Item Version Anxiety subscale (DASS-21; Antony & Bieling, 1998; Lovibond & Lovibond,

1995); the Positive and Negative Affect Schedule Fear subscale (PANAS; Watson et al., 1988; Watson & Clark, 1999); and the Overall Anxiety Severity and Impairment Scale (OASIS; Campbell-Sills et al., 2009; Norman et al., 2006). We use the Anxiety composite score to increase robustness of measuring anxiety as a construct. Because the questionnaires vary in their response scales (e.g., 0-3 for DASS-21, 1-5 for PANAS) and the number of items used in those subscales, we adjusted the scoring to scale them on a 0 to 1 scale. The following formula presents how we calculated our Anxiety composite score: ([(STAI Trait Anxiety – 1)*7/4] + [(DASS-21 Trait Anxiety)*7/4] + [(PANAS Fear – 1)*6/5] + [(OASIS)*5/5])/25. We preregistered the calculation of this score. Additionally, we measured the DASS-21 and OASIS twice: once during participant recruitment and once at the assessment session (prior to fear conditioning); the recruitment data was used to determine high/low anxiety and eligibility, whereas the assessment data was used in the calculation of our Anxiety composite score.

Trait Depression (i.e., "Depression").

Much like our composite Anxiety measure, we pre-registered the calculation of a trait Depression composite score using the STAI Depression subscale, DASS-21 Depression subscale, PANAS Sadness subscale, and PANAS Positive Affect subscale (reverse-coded). The calculation was ([(STAI Depression – 1)*13/4] + [(DASS-21 Depression)*7/4] + [(PANAS Sadness – 1)*5/5] + [(PANAS Positive Affect, Reverse-Coded – 1)*10/5])/35.

Procedure

Participants attended one experimental session where they provided informed consent, physiological equipment was attached, and a shock workup procedure was conducted. In the shock workup procedure, shocks started at a low intensity and increased to the level the participant considered "uncomfortable but not painful" using a 0-10 discomfort scale (0 = "Not at all," 5 = "Moderately," and 10 = "Very"; M = 5.737, SD = 1.392). Then, participants commenced the primary experimental phases: Habituation, Training, Reminder, and Transfer Test (see Table 1). The sequence of Training phases (i.e., differential conditioning, feature-positive training, feature-negative training) was counterbalanced between participants. After Training, participants engaged in the Reminder phase, which included all stimuli from Training and maintained their reinforcement schedule. Lastly, participants completed Transfer Test, which included two trials each of the putative positive occasion setter presented serially with the CS- (C→b) and the putative negative occasion setter presented serially with the CS+ (E→a). The US was not presented during Transfer Test.

Additionally, several studies have shown that simultaneous presentation of the occasion setter and CS leads both to form a direct association with the US (Holland, 1984, 1986a), whereas presenting the stimuli serially (i.e., occasion setter first, then the CS) with sufficient time gaps between each produces occasion setting (Holland, 1986a). Therefore, between our putative occasion setter and CS, there was an 8-sec trace interval that looked identical to an ITI (see Figure 1 for trial structure). During Training and Reminder, reinforcement rates were 100% for a+, C→d+, and f+ and 0% for b-, d-, and E→f-. No shocks were delivered during Habituation and Transfer Test. During Training and Reminder, trials were grouped into blocks of one reinforced trial and two non-reinforced trials. In occasion setting studies with non-human animals, having 2-4x more non-reinforced vs reinforced trials is common (Holland, 1984, 1986a, 1991b), and we

chose 2x in order to be consistent with previous literature, reduce the density/frequency of electric shocks, and maintain a relatively shorter experiment (compared to 3-4x non-reinforced stimuli). Within each block, trial sequence was randomized, ultimately resulting in pseudo-randomization with no more than two consecutive reinforced trials or four consecutive non-reinforced trials in a given phase. During Transfer Test, each compound was presented once (in randomized order) before being present a second time. The fear conditioning lasted approximately 70 minutes.

	Experiment 1 and 2 Training			
Experiment 1 Habituation	Differential Conditioning	Feature-Positive Training	Feature-Negative Training	Experiment 1 and 2 Reminder
1 each: **a** through **f**	8 **a+** 16 **b-**	8 **C→d+** 16 **d-**	8 **f+** 16 **E→f-**	2 each: **a+**, **C→d+**, **f+** 4 each: **b-**, **d-**, **E→f-**

	Experiment 2 Transfer Test			
Experiment 1 Transfer Test	Main Comparisons	Additional Comparisons		
2 **E→a-** 2 **C→b-**	2 **E→a** 2 **Ea** 2 **C→b** 2 **Cb** 2 **ab**	2 all trained stimuli 2 **C** 2 **E** 2 **G** 2 **H**	**a→b** **b→a** **G→a** **G→b** **Ga**	**Gb** **H→a** **H→b** **Ha** **Hb**

Table 1. Experiment 1 and 2 Design. The order of phases within the Training phase (i.e., Differential Conditioning, Feature-Positive Training, Feature-Negative Training) was counterbalanced across participants. During Training, stimuli were grouped into eight blocks of one reinforced stimulus with two non-reinforced stimuli (e.g., 1 a+, 2 b-). During Reminder, stimuli were grouped into two blocks that included one of each reinforced stimulus with two of each non-reinforced stimulus. During Transfer Test, stimuli were grouped into two blocks that included one of each stimulus. In all phases, stimulus order was randomized within each block. "+" indicates electric shock delivery; "-" indicates no electric shock delivery; "→" indicates serial presentation; two consecutive stimuli (e.g., Ea) indicates simultaneous presentation. For Experiment 2, G was a novel geometric shape, and H was a novel auditory sound.

Data Analysis

We used Stata 15.1 to conduct multilevel modeling for inferential statistics. All analyses were conducted per block of trials: during Training and Reminder, each block consisted of one reinforced and two non-reinforced trials (e.g., 1 C→d+, 2 d-); during Transfer Test, one block consisted of one presentation each of E→a and C→b. Level 1 predictors were Stimulus (a+, b-, C→d+, d-, E→f-, f+) and Trial (1, 2, etc.).

During Training, we modeled Trial as a continuous variable to assess effects across trials, allowing for linear and quadratic slopes. In the Results, the "quadratic model" refers to an intercept, linear slope, and quadratic slope; the "linear model" refers to an intercept and linear slope. If the quadratic slope was non-significant, we removed it and re-ran it as a linear model; if the linear slope was also non-significant, we removed it and re-ran it without Trial as a factor.

In the analysis of US expectancy and SCR during Transfer Test, we used the last block from Reminder (for trained stimuli: a+, b-, C→d+, d-, E→f-, f+) and the first block of Transfer Test (for transfer stimuli: E→a, C→b). To evaluate whether the putative occasion setters affected responding to the CS+ and CS- (therefore examining if the former were indeed occasion setters), we calculated the absolute differences between 1) E→a and a+, 2) E→a and E→f-, 3) C→b and b-, and 4) C→b and C→d+, where we used the larger mean value to subtract from the smaller mean value (e.g., C→d+ minus C→b). The goal of this was to determine whether the transfer stimuli were closer in value to the CS+/CS- than the trained compound stimuli.

In the analysis of self-report fear during Transfer Test, we compared fear ratings after Reminder (for trained stimuli: a+, b-, C→d+, d-, E→f-, f+) and after completion of Transfer Test (for transfer stimuli: E→a, C→b). Difference scores were calculated the same as in the previous paragraph.

For analyses involving Anxiety or Depression, each was included separately as a Level 2 predictor in analyses using the same structure as the above paragraphs. To provide more precise examination of simple effects, we then conducted block-by-block simple effects analyses to determine which specific blocks Anxiety or Depression had an effect on (e.g., earlier vs later blocks). We conducted these analyses without correction, as well as with Holm-Bonferroni correction (Holm, 1979) if significant results emerged, correcting for the number of simple effects tests in that analysis. Additionally, we checked for multicollinearity of our composite Anxiety and Depression measures in both studies using variance inflation factor (VIF) (Thompson et al., 2017), where scores of >10 are indicative of potential multicollinearity. Our VIF scores in Study 1 (1.71) and Study 2 (2.37) were well below this cutoff, suggesting we had no multicollinearity concerns.

Experiment 1 Results

Training and Reminder

The results across experimental phase are presented in Table 2 and Figure 2. For additional details, see Supplementary Materials Figures SM1 and SM4. Overall, results showed significant quadratic and linear effects for each measure of fear (US expectancy, self-report fear, SCR; all $\chi^2(5\text{-}7) > 56.43$, all $p < .001$), as well as greater responding for reinforced than non-reinforced trials within a+ vs b-, C→d+ vs d-, and E→f-, f+ (all $Z > 2.14$, all $p < .032$). Additionally, we assessed whether responding changed from the last block of Training to the last block of Reminder (US expectancy, SCR) or after Training to after Reminder (self-report fear). There were no significant changes with any stimulus for SCR (ps > .103) or self-report fear (ps > .515). For US expectancy, there were no significant changes with a+, b-, C→d+, or d- (ps > .056), but there was a significant movement towards uncertainty for E→f- ($Z = 2.50$, $p = .012$) and f+ ($Z = -3.63$, $p < .001$). We then analyzed E→f- vs f+ on the final Reminder block, finding that f+ had significantly greater US expectancy than E→f- ($Z = 11.09$, $p < .001$).

Table 2. Experiment 1 Training Results

Measure	Model	Simple Effects	χ^2	df	f	p	Z
US Expectancy	**Quadratic**		**257.58**	**7**	**.226**	**<.001**	
	Linear		**510.93**	**7**	**.319**	**<.001**	
		a+ vs b-				**<.001**	**12.10**
		C→d+, d-				**<.001**	**10.40**
		E→f-, f+				**<.001**	**-8.79**
		C vs E				.602	0.52
Self-Report Fear	**Quadratic**		**70.14**	**5**	**.175**	**<.001**	
	Linear		**152.08**	**5**	**.257**	**<.001**	
		a+ vs b-				**<.001**	**8.23**
		C→d+, d-				**.011**	**2.56**
		E→f-, f+				**<.001**	**-8.53**
SCR	**Quadratic**		**56.43**	**7**	**.106**	**<.001**	
	Linear		**86.80**	**7**	**.132**	**<.001**	
		a+ vs b-				**<.001**	**3.70**
		C→d+, d-				**.014**	**2.45**
		E→f-, f+				**.032**	**-2.14**
		C vs E				.420	0.81

Significant results in **bold**. Training went as expected: participants discriminated a+ vs b-, C→d+ vs d-, and E→f- vs f+ with all measures of fear.

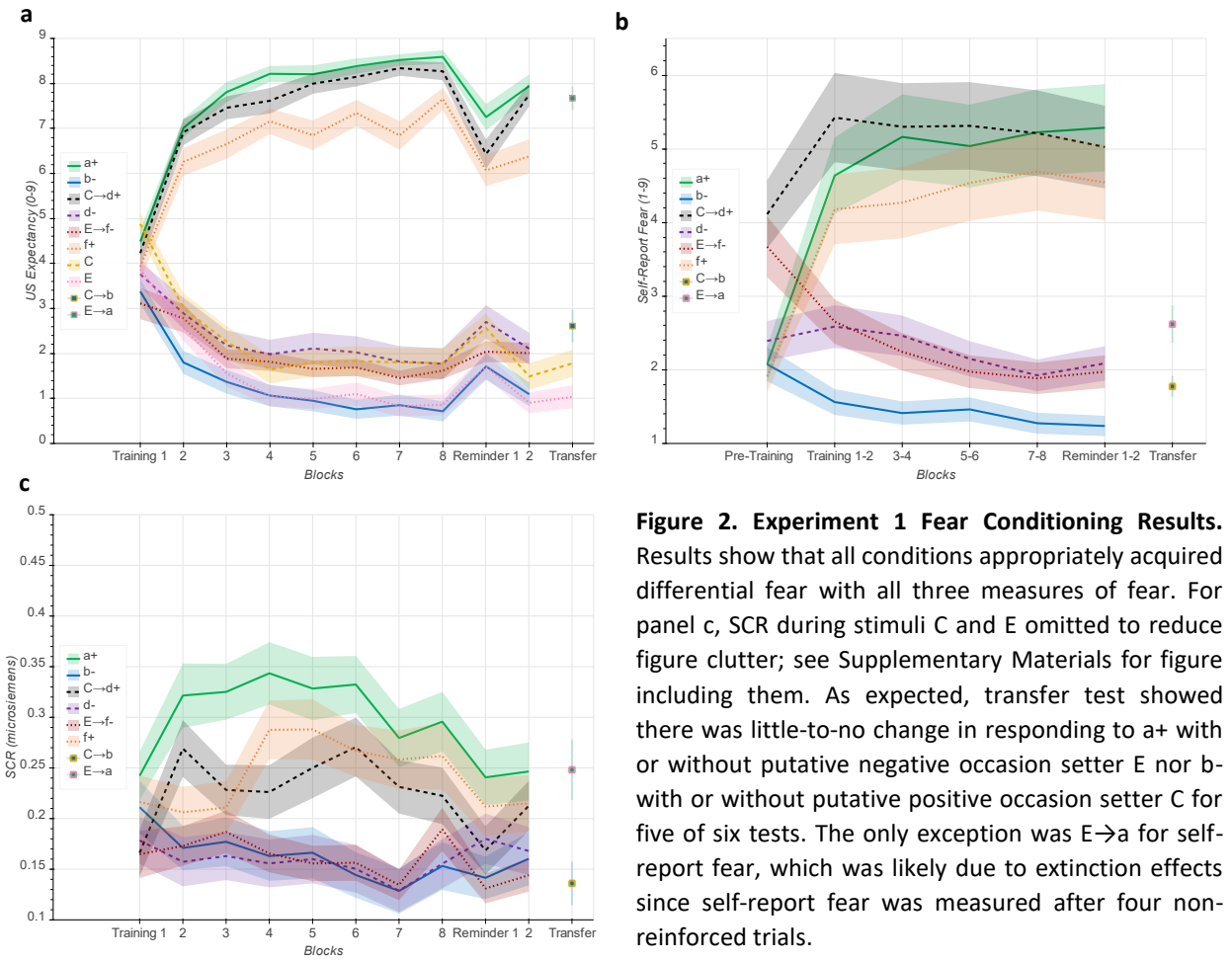

Figure 2. Experiment 1 Fear Conditioning Results. Results show that all conditions appropriately acquired differential fear with all three measures of fear. For panel c, SCR during stimuli C and E omitted to reduce figure clutter; see Supplementary Materials for figure including them. As expected, transfer test showed there was little-to-no change in responding to a+ with or without putative negative occasion setter E nor b- with or without putative positive occasion setter C for five of six tests. The only exception was E→a for self-report fear, which was likely due to extinction effects since self-report fear was measured after four non-reinforced trials.

Transfer Test

The results of Transfer Test are shown in Table 3 and Figure 2. We predicted the putative occasion setters (E and C) would not greatly affect responding to the CS+ and CS- (i.e., a+ and b-, respectively). This would be evidenced in the transfer test where E→a would produce more similar responding to a+ than E→f-, and C→b would produce more similar responding to b- than C→d+.

Overall, as expected, E→a was more similar to a+ than E→f- with US expectancy and SCR (all $t(77\text{-}78) < -2.617$, all $p < .011$). Additionally, as expected, C→b was more similar to b- than C→d+ with US expectancy, SCR, and self-report fear (all $t(78\text{-}79) > 3.806$, all $p < .001$). The only test which did not support the hypotheses was E→a with self-report fear ($t(78) = 4.427$, $p < .001$); however, this was likely due to extinction effects since self-report fear was measured after four non-reinforce trials during Transfer Test. Thus, with the exception of E→a self-report fear, the results were consistent with C and E being positive and negative occasion setters, respectively.

Table 3. Experiment 1 Transfer Test

Measure	Transfer Stimuli	Difference	SE	t	df	d	p
US Expectancy	**a+ minus E→a vs E→a minus E→f-**	**-5.520**	**0.552**	**-9.995**	**77**	**1.132**	**<.001**
	C→b minus b- vs C→d+ minus C→b	**3.667**	**0.711**	**5.159**	**78**	**0.580**	**<.001**
Self-Report Fear	**a+ minus E→a vs E→a minus E→f-**	**1.975**	**0.446**	**4.427**	**78**	**0.498**	**<.001**
	C→b minus b- vs C→d+ minus C→b	**2.713**	**0.392**	**6.915**	**79**	**0.774**	**<.001**
SCR	**E→a minus a+ vs E→a minus E→f-**	**-0.085**	**0.032**	**-2.617**	**78**	**0.299**	**0.011**
	b- minus C→b vs C→d+ minus C→b	**0.106**	**0.028**	**3.806**	**78**	**0.426**	**<.001**

Significant results in **bold**. We hypothesized that E→a would be closer to a+ than E→f-, and C→b would be closer to b- than C→d+. For five of six tests above, this hypothesis was supported (E→a for self-report fear was not).

Higher Trait Anxiety (But Not Trait Depression) is Associated with Greater Increases in Occasion Setting Fear during Middle-to-Late Blocks of Training.

See Figure 3 for effects of Anxiety on fear of each stimulus and Supplementary Materials Table SM1 for statistical details.

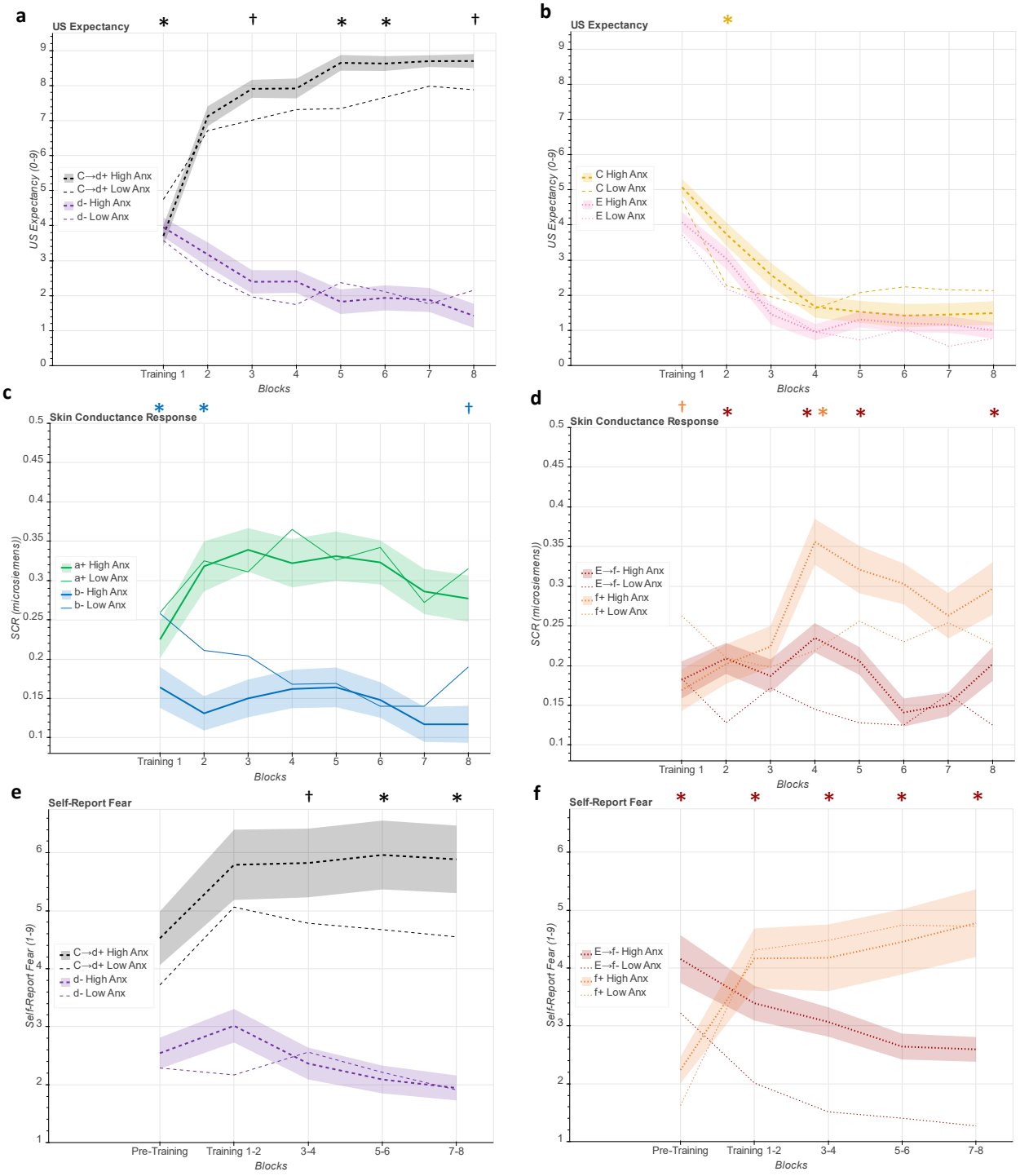

Figure 3. Experiment 1 Anxiety and Training Results. Only stimuli with significant effects shown to reduce figure clutter (see Supplementary Materials for remaining stimuli). Trait anxiety was analyzed as a continuous measure; for figure, anxiety was divided into high and low values for descriptive purposes (i.e., ± 1 SD from mean anxiety). "High" refers to +1 SD trait anxiety, and "Low" refers to -1 SD trait anxiety. Results show that High Anxiety predicted greater fear of C→d+ (US expectancy, self-report fear), E→f- (self-report fear, SCR), and f+ (SCR). Low Anxiety started with higher b- and had a greater decrease (SCR). Analysis of differential effects of Anxiety on fear (e.g., Anxiety on C→d+ vs d-) showed Anxiety had a stronger effect of increasing fear to C→d+ than d- (US expectancy, self-report fear) and E→f- than f+ (self-report fear). Error bands represent standard error and are identical within each stimulus for High and Low Anxiety; for visual clarity, error bands only placed on High Anxiety. Uncorrected block-by-block effects of Anxiety: * $p < .05$, † $p < .10$.

First, higher Anxiety predicted greater increases in fear to C→d+ (US expectancy, self-report fear), E→f- (self-report fear, SCR), and f+ (SCR), where the most consistent finding was greater fear on later blocks for high vs low Anxiety. Specifically, the Anxiety x Stimulus x Trial quadratic model was not significant with any measure of fear (all $\chi^2(5\text{-}7) < 13.25$, all $p < .066$), but the Anxiety x Stimulus x Trial linear model was significant for US expectancy and SCR (all $\chi^2(7) > 16.18$, all $p < .024$), and the Anxiety x Stimulus interaction was significant for self-report fear ($\chi^2(5) = 35.92$, $p < .001$). Uncorrected block-by-block simple effects showed higher Anxiety predicted significantly lower Block 1 fear for C→d+ (US expectancy; $Z = -2.18$, $p = .029$) and marginally lower Block 1 fear for f+ (SCR; $Z = -1.71$, $p = .088$). Additionally, higher Anxiety predicted higher fear on later blocks for C→d+ (US expectancy, self-report fear), E→f- (self-report fear, SCR), f+ (SCR), and during occasion setter C (US Expectancy). Specifically, higher Anxiety predicted marginally (Blocks 3 and 8: ps = .060 and .086) and significantly (Blocks 5 and 6: ps = .006 and .044) greater C→d+ US expectancy, marginally (Blocks 3-4; $p = .077$) and significantly (Blocks 5-6 and 7-8; ps < .028) greater C→d+ self-report fear, significantly greater E→f- self-report fear (all Blocks: ps < .016), significantly greater E→f- SCR (Blocks 2, 4, 5, and 8: ps < .046), significantly greater f+ SCR (Block 4: $p = .012$), and significantly greater US expectancy during occasion setter C (Block 2: $p = .022$).

Furthermore, we conducted analyses assessing the effects of Anxiety within training discriminations in which there was a significant solo effect (e.g., a+ vs b-, C→d+ vs d-, and E→f- vs f+ if one of the stimuli within each pair had a significant solo effect from previous paragraph). The Anxiety x Stimulus x Trial linear model was significant for US expectancy C→d+ vs d- ($\chi^2(1) = 8.94$, $p = .004$) and marginally significant for SCR E→f- vs f+ ($\chi^2(1) = 3.76$, $p = .053$), and the Anxiety x Stimulus interaction was significant for self-report fear E→f-

vs f+ and C→d+ vs d- (all $\chi^2(1) > 10.01$, all $p < .002$). Uncorrected block-by-block effects showed that higher Anxiety was associated with higher responding to d- than C→d+ on Block 1 for US expectancy ($Z = -2.01$, $p = .045$) but higher responding to C→d+ than d- for US expectancy Blocks 5 and 8 ($Zs = 2.61$ and 2.19, $ps = .009$ and $.029$) and self-report fear Blocks 3-8 ($Zs \geq 2.01$, $ps \leq .045$). Also, block-by-block effects showed that higher Anxiety was associated with greater self-report fear for E→f- than f+ after Blocks 1-2, 3-4, 5-6, and 7-8 ($ps < .030$) but not at Pre-Training ($p = .585$), and there was no significant effect of E→f- vs f+ for SCR on any blocks ($ps > .103$). Thus, the effects of Anxiety showed consistent effects when assessing stimuli/compounds alone (e.g., C→d-) and differentially for C→d+ vs d- (US expectancy, self-report fear) and E→f- vs f+ (self-report fear but not SCR). This suggests that higher Anxiety was associated with greater differential fear of the compound trial types (C→d+, E→f-) than singular trial types (d-, f+).

Second, there were some instances in which Anxiety was associated with lower fear. Specifically, uncorrected analyses showed that lower Anxiety predicted significantly greater SCR on Blocks 1 and 2 ($Zs = -2.43$ and -2.06, $ps = .015$ and $.039$) and marginally greater SCR on Block 8 ($Z = -1.90$, $p = .058$). There were no other significant block-by-block effects across measures of fear. We then assessed the association between Anxiety and differential responding to b- vs a+ with SCR. There was no significant Anxiety x Stimulus x Trial interaction for the quadratic ($\chi^2(1) = 1.08$, $p = .300$) or linear ($\chi^2(1) = 1.52$, $p = .217$) models, but there was a marginally significant Anxiety x Stimulus interaction ($\chi^2(1) = 3.45$, $p = .063$). For completeness, we examined block-by-block effects, finding no significant differential effects of b- vs a+ on any block ($ps > .136$). This suggests that lower Anxiety predicts greater SCR on the first two blocks for b- when assessed alone, but there was no differential effect of Anxiety on b- vs a+. Lastly, higher Anxiety was

associated with lower slope for d- and E→f- with US expectancy, but there were no significant block-by-block effects, limiting interpretability. Thus, while some significant interactions occurred in which lower Anxiety predicted higher fear, this was limited to early blocks (b- SCR) or had no block-by-block effects (d- and E→f- US expectancy), therefore lacking clarity or robustness.

Furthermore, we followed up the Anxiety analyses above with Holm-Bonferroni correction for the number of block-by-block analyses conducted. This provides a more conservative analysis of significant effects. Ultimately, surviving significant analyses showed that higher Anxiety predicted greater increases in fear of compound occasion setting stimuli. Specifically, higher Anxiety predicted higher US expectancy for C→d+ Block 5 ($p = .006$, cutoff = .00625); self-report fear for E→f- at Pre-Training ($p = .016$, cutoff = .05), Blocks 1-2 ($p = .001$, cutoff = .0125), 3-4 ($p < .001$, cutoff = .01), 5-6 ($p = .002$, cutoff = .025), and 7-8 ($p = .001$, cutoff = .0167); and differential E→f- vs f+ self-report fear for Blocks 1-2 ($p = .009$, cutoff = .0167), 3-4 ($p = .1$ cutoff = .01), and 5-6 ($p = .009$, cutoff = .0125). Thus, while the corrected analyses provide more conservative results, they are consistent with the most robust uncorrected results showing that higher Anxiety is predictive of greater increases in compound occasion setting stimuli (C→d+, E→f-).

Lastly, for specificity, we examined the effects of trait Depression, finding no significant results in any model for US expectancy ($ps > .055$) but a significant Depression x Stimulus interaction for self-report fear ($\chi^2(5) = 17.68$, $p = .003$) and a significant quadratic ($\chi^2(7) = 16.10$, $p = .024$) and linear ($\chi^2(7) = 15.83$, $p = .027$) effect of Depression for SCR. Uncorrected simple effects for self-report fear were not significant. Uncorrected simple effects for SCR showed that higher Depression predicted a linear increase ($Z = 3.06$, $p = .002$) and quadratic

deceleration/decrease (Z = -2.69, p = .007) for b- across trials, as well as a linear decrease (Z = -2.29, p = .002) and quadratic increase/acceleration (Z = 2.70, p = .007) for C→d+ across trials. Block-by-block analyses showed only a marginal effect of Depression for b- on Block 6 (Z = 1.85, p = .065) and significantly greater SCR for higher Depression with C→d+ Blocks 7 (Z = 2.30, p = .021) and 8 (Z = 2.09, p = .037). However, none of these uncorrected effects survived Holm-Bonferroni correction. Lastly, the effects of Depression on differential US expectancy for C→d+ vs d- were significant for both the quadratic ($\chi^2(1) = 5.15$, p = .023) and linear ($\chi^2(1) = 5.25$, p = .022) models. However, simple effects showed no block-by-block effects (ps > .085). Thus, in total, Depression showed some scattered associations with self-report fear and SCR on some stimuli, but block-by-block effects were largely non-significant, and those that were did not survive Holm-Bonferroni correction, ultimately suggesting Depression had no reliable effect on fear conditioning Training.

Higher Trait Anxiety is Marginally Associated with Greater Self-Report Fear of Transfer Test Stimuli E→a and C→b. No Effects of Trait Depression.

We assessed whether Anxiety was predictive of Transfer Test stimuli E→a and C→b. Results showed that there was no Anxiety x Stimulus interaction nor a main effect of Anxiety with US expectancy (ps > .244) or SCR (ps > .323). The Anxiety x Stimulus interaction was not significant with self-report fear (p = .975), though the main effect of Anxiety was marginally significant (p = .051, b = 3.061, SE = 1.571) with higher Anxiety predicting greater self-report fear. For completeness, we examined E→a and C→b separately, finding no significant effects of either compound (ps = .116 and .125, respectively). Additionally, there were no effects of

Depression with any model or any measure of fear (ps > .100). See Supplemental Materials Table SM2 for statistical details.

Experiment 1 Discussion

This was the first study to investigate human Pavlovian occasion setting using traditional fear conditioning methods with tests of whether the underlying learning was indeed occasion setting (i.e., transfer tests). We tested three forms of learning: differential conditioning (a+, b-), feature-positive training (C→d+, d-), and feature-negative training (E→f-, f+). Results showed that participants acquired greater fear of stimuli associated with electric shock than those associated with the absence of shock. Furthermore, individuals with high (vs low) trait anxiety largely demonstrated greater fear of CSs trained with occasion setters (i.e., C→d+, E→f-) during Training, whereas trait depression had little-to-no effect. Additionally, our transfer tests suggested that occasion setting was learned in the feature-positive and feature-negative conditions, as evidenced by little-to-no transfer of the putative occasion setters to the CS+ and CS-.

Experiment 2

Introduction

Although our transfer test from Experiment 1 was consistent with occasion setting, we did not evaluate presence of transfer or summation where expected (e.g., presenting a+ and b- in compound) to show whether a+ and b- could be acted upon in a novel combination (Laing et al., 2021). Thus, we conducted a second experiment (appetitive conditioning) using many of the same procedures as in the fear conditioning for comparison of results between studies. There were two primary aims: conduct more transfer tests to determine whether occasion setting was learned, and assess the association between trait anxiety and trait depression and occasion setting learning in the appetitive domain. We additionally investigated the effects of trait depression because of depression's association with poorer reward processing (Admon & Pizzagalli, 2015). In conjunction with Experiment 1, this allowed us to assess specificity of effects between 1) trait anxiety and trait depression, 2) fear conditioning vs appetitive conditioning, and 3) direct associative learning vs occasion setting.

Methods

Participants

Ninety participants were collected online using Prolific (www.prolific.co). As pre-registered (https://osf.io/3au7v), eight participants were removed due to inattention (e.g., invariant responding, failure to discriminate a+ from b-), leading to a final sample size of N = 82. Participants were 60.98% female, 36.59% male, 1.22% female-to-male transgender, and 1.22% male-to-female transgender; mean age 30.76 years (SD = 10.40); and 7.32% Black or African-

American, 10.98% Central/East Asian, 7.32% Hispanic/Latinx, 57.32% White, and 7.32% South Asian, and 9.76% multiracial. Using the same trait anxiety cutoffs as in Experiment 1, there were relatively more participants with high trait anxiety in Experiment 2 than Experiment 1 (Experiment 1: 40 High, 40 Low; Experiment 2: 53 High, 29 Low). Participants were paid $13.30 for completing the ~60 min experiment. This study was deemed exempt by the California Institute of Technology Institutional Review Board, and all participants provided informed consent prior to commencing the study.

Materials and Apparatus

All materials were the same as in Experiment 1 except where noted below. We programmed the experiment using PsychoPy version 2020.1.3 (Peirce et al., 2019). We used the same geometric shapes for CS images as in Experiment 1 (green star, blue triangle, black square, white circle), as well as an additional visual stimulus (stimulus G; purple curved parallelogram), which was used as a novel visual stimulus during transfer test. All visual stimuli were counterbalanced via a Latin square (5 versions). We used the same auditory stimuli as in Experiment 1 (violin, trumpet), as well as a novel white noise stimulus (stimulus H). All auditory stimuli were counterbalanced via a Latin square (3 versions). As in Experiment 1, the sequence of the three Training phases (differential conditioning, feature-positive training, feature-negative training) was fully counterbalanced (6 versions). Thus, we created 90 versions of the experiment multiplying the counterbalancing versions above. Participants were then randomized to one of these versions. The US was monetary reward ($0.30), which was an audio-visual stimulus (picture of a 30¢ gold coin with confetti and audio of cash register sound (“cha-ching!”)). During Transfer

Test, the US was muted and covered with an image of a curtain to prevent learning about (non)reinforcement (e.g., to prevent extinction effects). Participants were told that, while they would not know whether they got the monetary reward during a given trial, they would still receive the payment at the end of the experiment if the trial resulted in the monetary reward. Participants passed a quiz demonstrating their understanding of the curtain. Thus, we were able to test many stimuli during Transfer Test without extinction effects. ITIs were reduced to 1.25 seconds to reduce study duration and maintain participant engagement; trials ended with a US expectancy rating.

US expectancy.

At the end of every trial, participants rated US expectancy using a visual analog scale: "How certain are you that you are about to receive a bonus payment?" There were no numerical anchors shown on the visual analog scale, but the verbal anchors at the left, middle, and right were "Certain no bonus," "Completely uncertain," and "Certain yes bonus." Results were transformed to a 0-9 scale to match the US expectancy scale from Experiment 1.

Procedure

Participants completed trait questionnaires as in Experiment 1. Then, they commenced the conditioning: Training, Reminder, and Transfer Test. We removed the Habituation phase because we were not measuring SCR and did not need this phase to reduce physiological responding due to stimulus novelty. During Transfer Test, we included many stimuli (see Table 1). The most critical comparisons were C→b and E→a vs ab ("→" indicates serial presentation with an 8sec

trace, whereas two stimuli together (e.g., ab) indicates simultaneous stimulus presentation). We predicted C→b < ab < E→a. We also tested Cb and Ea and predicted Cb < ab < Ea. Additionally, we tested all of the trained stimuli in the arrangements in which they were trained (e.g., a, b, C→d, d, E→f, f). We expected a change in US expectancy with novel arrangements compared to trained arrangements due to the former's novelty (where novel arrangements would produce more "uncertain" ratings closer to a value of 4.5 US expectancy). Thus, to account for this novelty-induced responding, we included several control comparisons. Using both serial and simultaneous compounds, we tested G (a novel geometric shape) and H (a novel auditory stimulus) in place of C and E (i.e., Ga, Gb, G→a, G→b; Ha, Hb, H→a, H→b). For completeness, we also tested the effects of one trained CS presented serially with the other trained CS (i.e., a→b, b→a). We generally predicted C and E would produce similar responding as G and H to targets a and b.

Data Analysis

Analyses were conducted in a similar manner as Experiment 1. Analyses for Transfer Test differed in Experiment 2 due to different experimental methods. In Experiment 1, because there was no central comparison stimulus combination (e.g., ab), we calculated difference scores to observe if 1) E→a was more similar to a+ than E→f-, and 2) C→b was more similar to b- than C→d+. However, in Experiment 2, we tested ab, so our comparisons stemmed from this. We analyzed our main comparisons of b, C→b, Cb, ab, E→a, Ea, and a, with the most critical prediction being C→b and Cb < ab < E→a and Ea.

Experiment 2 Results

Training and Reminder

See Figure 4 for results. Overall, results for Training showed significant quadratic ($\chi^2(5) = 282.96$, $f = .271$, $p < .001$) and linear ($\chi^2(5) = 458.38$, $f = .345$, $p < .001$) effects for US expectancy, as well as greater responding for reinforced than non-reinforced trials within a+ vs b- ($Z = 13.13$, $p < .001$), C→d+ vs d- ($Z = 11.59$, $p < .001$), and E→f-, f+ ($Z = -12.22$, $p < .001$). Additionally, we assessed whether responding changed from the last block of Training to the last block of Reminder, finding no significant changes for a+ ($p = .191$), b- ($p = .133$), E→f- ($p = .292$), or f+ ($p = .172$), but there was a significant movement towards uncertainty for C→d+ ($Z = -2.18$, $p = .030$) and d- ($Z = 2.72$, $p = .006$). However, on the final Reminder block, there was significantly greater US expectancy for C→d+ than d- ($Z = 18.06$, $p < .001$).

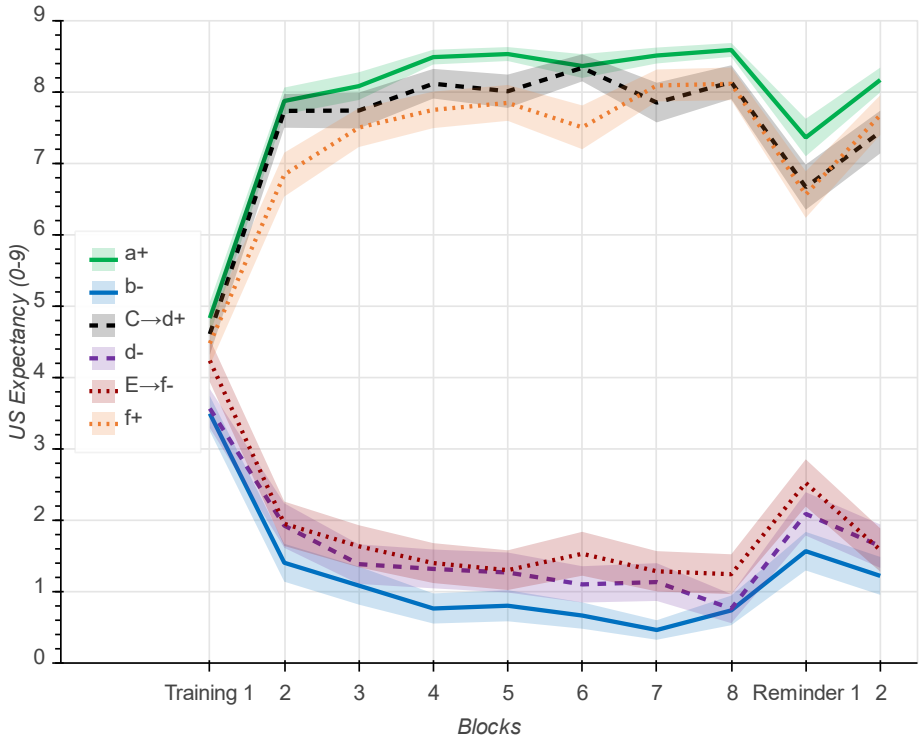

Figure 4. Experiment 2 Appetitive Conditioning Results. Results show that all conditions appropriately acquired differential responding. Error bands show standard error. Results are highly similar to the US expectancy results from Experiment 1 (see Figure 2a), suggesting training was similarly effective across both experiments.

Transfer Test

The results of Transfer Test are shown in Figure 5, and statistical details are in Supplementary Materials Table SM3. Overall, as expected, the putative positive occasion setter and CS- (C→b and Cb) had lower US expectancy than the compound CS+ and CS- (ab) (Zs > 2.71, ps < .007). Similarly, as expected, the putative negative occasion setter and CS+ (E→a and Ea) had higher US expectancy than the compound CS+ and CS- (ab) (Zs > 2.53, ps < .011). Each of the above simple effects survived Holm-Bonferroni correction, suggesting their reliability. C→b and Cb did not significantly differ (Z = .04, p = .972), nor did E→a and Ea (Z = .23, p = .820), suggesting that serial or simultaneous presentation made no difference in the effect of the putative occasion setters on CS+ or CS- responding.

Additionally, there was 1) greater responding to C→b and Cb than b- (Zs > 6.04, ps < .001), and 2) lower responding to E→a and Ea than a+ (Zs > 6.63, ps < .001). These effects survived Holm-Bonferroni correction. This could be due to transfer effects or novelty of the compound. To assess this, we compared responding to a and b when combined with 1) trained putative occasion setters C and E, and 2) novel stimuli G and H (see Supplementary Materials Table SM3 and Figure SM5 for details). In short, responding to a and b when combined with these stimuli was largely similar. Out of eight tests, six showed no significant differences as expected; the two differences that emerged were Cb > Hb (Z = 2.05, p = .041) and E→a < G→a (Z = -2.30, p = .021). However, these significant results did not survive Holm-Bonferroni correction (i.e., Cb vs Hb: cutoff = .00714; E→a vs G→a: cutoff = .00625), suggesting these effects were not reliable. This demonstrates that novelty of the stimulus combinations was responsible for the increases in responding to b and decreases in responding to a when combined with C, E, G, or H. Thus, it

appears that the underlying learning that occurred with C and E had little-to-no effect on responding to a and b, providing support that they were indeed occasion setters.

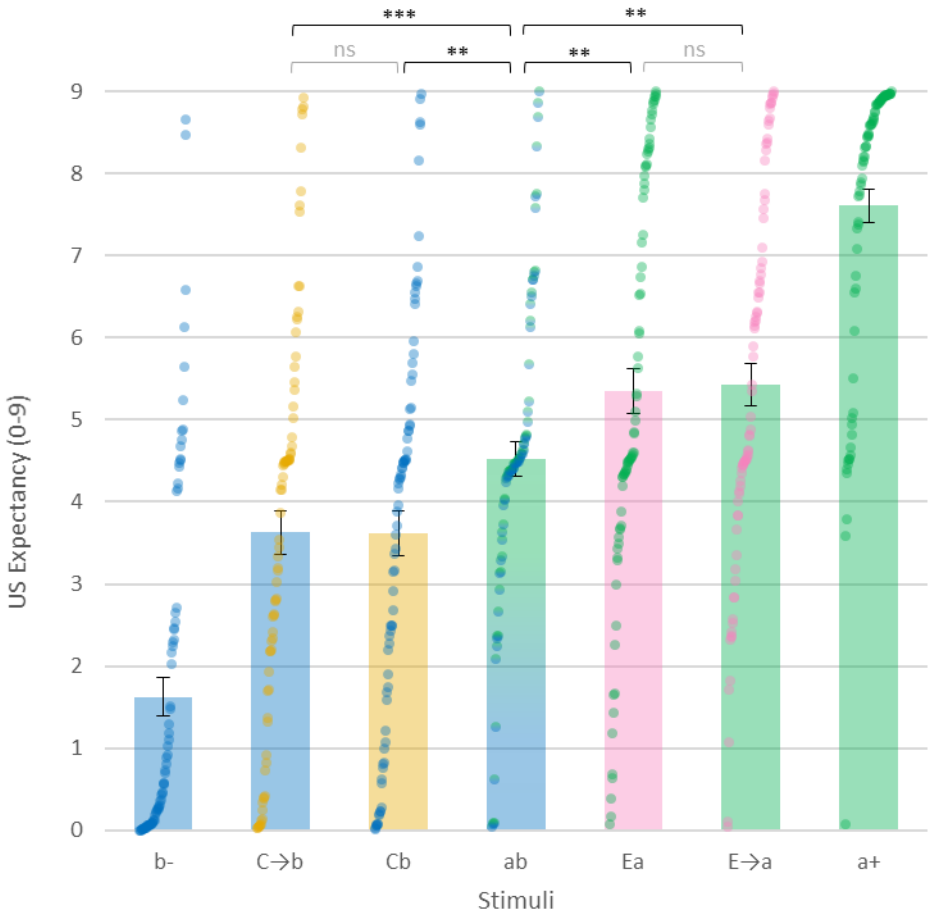

Figure 5. Experiment 2 Transfer Test. Bars indicate mean, error bars indicate standard error, and dots indicate individual data points arranged in ascending order within stimulus per an empirical cumulative distribution function. Primary hypotheses were C→b and Cb < ab < Ea and E→a; results supported these hypotheses. b- < all other stimuli; a+ > all other stimuli. For further stimulus comparisons, see Supplemental Materials Figure SM5.

Trait Anxiety and Trait Depression Not Reliably Associated with Stimuli during Training

There were no significant effects of Anxiety (ps > .192) in any model or Depression (ps > .637) in most models during Training. The one exception was a main effect of Depression, where greater Depression predicted lower US expectancy ($Z = -2.26$, $p = .024$). We examined simple effects per stimulus, finding that greater Depression predicted lower US expectancy for b- ($Z = -2.41$, $p = .016$) and d- ($Z = -2.16$, $p = .031$) but not the remaining stimuli (ps > .142). However, these did not survive Holm-Bonferroni correction (lowest cutoff = .00833).

Higher Trait Depression (But Not Trait Anxiety) Associated with Lower Monetary Expectancy for Trained Negative Occasion Setting Compound and Negative Occasion Setter/CS+ Serial Compound

There was no Anxiety x Stimulus interaction ($\chi^2(24) = 21.07$, $p = .635$) nor a main effect of Anxiety ($Z = -1.70$, $p = .090$) when assessing all transfer stimuli together in one test. For completeness, we examined simple effects per stimulus, finding that higher Anxiety was associated with lower US expectancy for Ea ($Z = -2.88$, $p = .004$), E→a ($Z = -1.96$, $p < .050$), and Ha ($Z = -2.51$, $p = .012$), with marginal effects for H→a ($Z = -1.80$, $p = .072$). However, these results did not survive Holm-Bonferroni correction.

Furthermore, there was a significant Depression x Stimulus interaction ($\chi^2(24) = 40.58$, $p = .019$). Results showed that higher Depression was associated with lower US expectancy for many stimuli, including trained/non-reinforced stimuli (b- and E→f-), as well as untrained/novel stimuli/compounds: a→b, Gb, ab, Ea, E→a, Ha, H→a, and H (Zs < -2.11 ps < .035), with marginal

effects for G→b (Z = -1.88, p = .060) and H→b (Z = -1.72, p = .086). Holm-Bonferroni-corrected analyses controlling for all transfer test stimuli found that E→a (p = .001, cutoff = .002) and E→f- (p = .002, cutoff = .00208) survived the correction, where higher Depression predicted lower monetary expectancy.

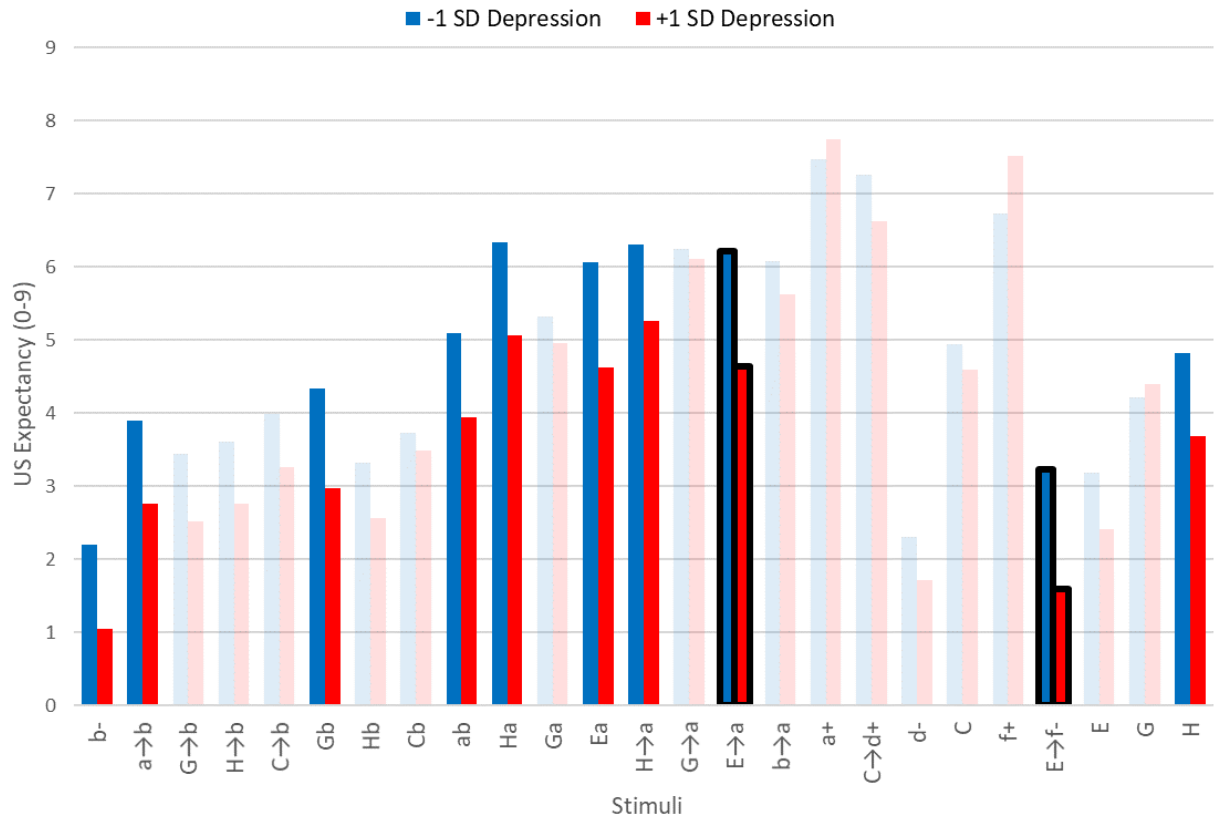

Figure 6. Experiment 2 Trait Depression and Transfer Test. Bars indicate mean values for trait Depression ± 1 SD from the mean. Stimuli with uncorrected significant effects are shown in bold colors; stimuli with significant effects that survive Holm-Bonferroni correction additionally have black borders. Corrected results show that trait depression was associated with lower monetary expectancy for the trained negative occasion setting compound (E→f-) and transfer of the negative occasion setter to an otherwise consistently reinforced CS+ (E→a).

Experiment 2 Discussion

Training results in this experiment were similar to the US expectancy results of Experiment 1, where participants learned which stimuli were (non)reinforced. Importantly, transfer test was consistent with occasion setting. Participants showed lower US expectancy to both serial and simultaneous combinations of the positive occasion setter and CS- (C→b, Cb) compared to simultaneous presentation of the CS+ and CS- (ab), suggesting that C was a positive occasion setter. Similarly, participants showed greater US expectancy to both serial and simultaneous combinations of the negative occasion setter and CS+ (E→a, Ea) compared to simultaneous presentation of the CS+ and CS- (ab), suggesting that E was a negative occasion setter. This showed that a and b were able to be acted upon in novel combinations, but the effects of C and E on b and a, respectively, were more muted than would be expected if C and E were simply CSs with direct excitatory or inhibitory associations with the US. Furthermore, we examined the effects of trait anxiety and trait depression on US expectancy during training and transfer test. There were several results that did not survive Holm-Bonferroni correction (see Results). The tests that did survive correction showed that higher trait depression during transfer test was associated with reduced expectancy for monetary reward associated with compound negative occasion setting – including the trained compound (E→f-) and the untrained, novel negative occasion setter/CS+ compound (E→a). The latter suggests that individuals high vs low in trait depression transferred the negative occasion setter more strongly to the otherwise 100% reinforced CS+.

General Discussion

Experiment 1 was fear conditioning with an electric shock US, and Experiment 2 was appetitive conditioning using a monetary reward US. In both experiments, training results showed that participants appropriately learned which stimuli were reinforced or not reinforced. Additionally, transfer test results in both experiments were consistent with occasion setting. Specifically, in Experiment 1, the putative negative occasion setter E had no effects on US expectancy and SCR during the CS+ (a+). Similarly, there were minor effects of putative positive occasion setter C on US expectancy and SCR during the CS- (b-); with US expectancy, there was a modest increase with C→b vs b-, and with SCR, there was a modest decrease with C→b vs b-. These modest changes often occur in occasion setting experiments, likely due to generalization between stimuli and novelty of the compounds (Laing et al., 2021), and are thus unlikely to be meaningful. Our Experiment 1 transfer test with self-report fear is likely less valid than US expectancy or SCR because of extinction effects; self-report fear during transfer test was measured after four non-reinforced trials, likely causing a methodological issue and reducing E→a responding. Thus, US expectancy and SCR comprise our most valid tests and are consistent with C and E as being occasion setters.

However, one limitation of Experiment 1's transfer test was not demonstrating that a+ and b- can be acted upon by other stimuli in novel compounds. Thus, in Experiment 2, we conducted transfer tests with a variety of other stimuli, including novel auditory and visual stimuli to match the trained stimuli in physical properties. Importantly, we observed that the putative positive occasion setter and CS- compound (i.e., serially presented: C→b, simultaneously presented: Cb) had lower US expectancy than the CS+ and CS- compound (i.e., ab). Similarly, the putative

negative occasion setter and CS+ compound (i.e., E→a, Ea) had greater US expectancy than ab. This suggests that a and b were able to be acted upon, as demonstrated by their summation, and that the effects of the putative occasion setters were less than the direct CSs upon each other. Additionally, the putative occasion setters had similar effects on a and b as novel, untrained stimuli G and H that were matched in physical properties. Thus, the evidence suggests that C and E were positive and negative occasion setters, respectively, in Experiment 2. Since many of the methods were matched across experiments, Experiment 2's results may provide additional support to Experiment 1's transfer tests, which were also consistent with C and E being occasion setters.

To our knowledge, this is the first report to investigate the effects of 1) trait anxiety and trait depression on 2) positive and negative occasion setting and with discrete stimuli 3) in both fear conditioning and appetitive conditioning. Our Holm-Bonferroni corrected fear conditioning results showed that higher trait anxiety was associated with greater increases in fear of compound stimuli from positive and negative occasion setting during training (i.e., C→d+, E→f-, and f+); there were no effects of trait depression. Our appetitive conditioning results showed the opposite: trait anxiety had no effects, whereas higher trait depression during transfer test showed lower expectancy of monetary reward for the trained negative occasion setting compound (E→f-) and transfer of the negative occasion setter to the otherwise 100% reinforced CS+ (E→a). This suggests specificity: trait anxiety (but not trait depression) is associated with faster learning/increases in fear conditioning with occasion setting compounds, and trait depression (but not trait anxiety) is associated with lower expectancy of monetary reward with the trained negative occasion setter/CS compound and transfer of the negative occasion setter to the CS+.

The trait anxiety/fear conditioning results extend findings that anxious individuals have greater fear with context-dependent extinction learning, which is believed to be a form of negative

occasion setting (Acheson et al., 2012; Barrett & Armony, 2009; Haaker et al., 2015; Liberzon & Sripada, 2007; Maren et al., 2013; Staples-Bradley et al., 2018; Trask et al., 2017). Our results also extend meta-analytic findings (Duits et al., 2015; Lissek et al., 2005), which showed that individuals with anxiety disorders have elevated responding to both dangerous and safe stimuli when learning direct associations (i.e., a+, b-) and during extinction of the dangerous stimulus (i.e., a+). While our study did not specifically replicate meta-analytic findings for simple stimuli (i.e., a+ and b-), it conceptually replicated the finding that higher trait anxiety predicts higher fear and extended those findings to occasion setting. One possible reason for the simple stimulus discrepancy in our study vs the meta-analyses is that the effects of trait anxiety on fear may be larger in occasion setting than direct associations, which allowed us to detect our occasion setting effects. Indeed, anxious individuals are more fearful in the presence of uncertainty than certainty or compared to non-anxious individuals (Carleton et al., 2012; Chan & Lovibond, 1996; Jensen et al., 2016; Wong & Lovibond, 2018). Because the CSs from occasion setting may have more uncertainty/ambiguity than simple CS+s/CS-s (due to the former's mixed association with the US), this may explain why we found effects of trait anxiety in occasion setting. Also, the meta-analyses compared clinically anxious individuals with healthy controls, whereas our study recruited participants based on high/low trait anxiety levels, which may have led to our difference in results for simple stimuli.

Furthermore, the trait depression/appetitive conditioning results are consistent with other research showing that depression is associated with deficits in reward processing (Admon & Pizzagalli, 2015). Specifically, depression is associated with avoidance of potentially rewarding environments (Smoski et al., 2008) and deficient processing of positive outcomes (Amsterdam et al., 1987; Branco et al., 2017; Potts et al., 1997; Schaefer et al., 2010; Scinska et al., 2004; Starr &

Hershenberg, 2017; Wu et al., 2017). Modern theories of reward processing state that reward processing is composed of three stages: motivation/anticipation of reward (i.e., "wanting" reward), consumption of reward (i.e., "liking" reward), and reward learning (e.g., learning which behaviors lead to reward; Berridge et al., 2009; Castro & Berridge, 2014; Craske et al., 2016; Rømer Thomsen et al., 2015). Depression is associated with deficits in the motivation/anticipation of reward; for example, individuals with depression have difficulty imagining future positive events (Holmes et al., 2009) and expend less effort to gain potentially higher reward (Treadway et al., 2012). Depression is also associated with deficits in reward learning (Kumar et al., 2008, 2018; Vrieze et al., 2013). However, the evidence linking depression with deficits in reward consumption (i.e., "liking" of reward) is mixed. One explanation for the mixed results is that anhedonia, rather than depression, is specifically is associated with reduced "liking" of reward (Berlin et al., 1998; Clepce et al., 2010; Keedwell et al., 2005; Wacker et al., 2009). In support, anhedonia has been associated with reduced "liking" of reward above and beyond the negative symptoms of anxiety and depression (Wacker et al., 2009). Our results here contribute to the depression/reward processing literature with most relevance to the reward learning phase, showing that individuals with depression in a learning task have more accurate expectations for the non-rewarded negative occasion setting compound stimuli (consistent with depressive realism; Moore & Fresco, 2012) and lower expectations for the novel and ambiguous negative occasion setter/CS+ compound. This may suggest that depressed individuals expect less reward in negative occasion setting compounds and their ambiguous transfer to an otherwise 100% rewarded stimulus (the CS+).

Furthermore, these experiments are highly relevant for anxiety and mood disorders. Trait anxiety and depression are associated with increased chances of clinically severe forms of each emotional category (Chambers et al., 2004; Mundy et al., 2015; Zinbarg et al., 2016). Our

experiments showed that individuals high vs low in trait anxiety demonstrated greater increases in fear in occasion setting, which suggests occasion setters might be an important treatment target in exposure therapy. For example, one potential improvement to exposure therapy could be extinction of the positive-occasion-setting abilities of positive occasion setters (i.e., their ability to modulate a CS/US association). There have been only a handful of experiments demonstrating extinction of occasion setting in humans (Franssen et al., 2017; van Vooren et al., 2012) and animals (Miller & Oberling, 1998; Rescorla, 1986), but extinguishing the modulatory properties of positive occasion setters in conjunction with extinguishing CSs could be beneficial. Additionally, it could be helpful to generate multiple negative occasion setters so that they generalize across situations and signal that the CS will not result in the US, similar to conducting exposure in multiple contexts (Craske et al., 2014). Generating discrete stimulus negative occasion setters may also be helpful much like conducting exposures in multiple contexts. Some specific examples of discrete stimulus occasion setters that could be relevant in clinical practice include: 1) a charismatic speaker gives their speech (positive occasion setter), then client expects their own speech (CS) will likely result in rejection (US), 2) an unskilled speaker gives their speech (negative occasion setter), then the client expects their own speech (CS) is unlikely to elicit rejection (US), 3) if a dog recently played with another dog (positive occasion setter), the client may worry that the dog (CS) is riled up and may attack the client (US), 4) if the dog was recently given a treat (negative occasion setter), the client may believe the dog (CS) is less likely to attack the client (US). Regarding trait depression and mood disorders, our results suggest that depressed individuals may strongly learn negative occasion setting non-reinforcement and may under-expect reward when the negative occasion setter is combined with an otherwise consistently rewarded CS+. For example, depressed individuals might learn that a rainy day (negative occasion setter) leads to less enjoyment (US) of playing outdoor

sports (CS), but the rainy day might also lead to less enjoyment of an otherwise consistently enjoyable activity (e.g., spending time with friends and family). Our results suggest that further research on the association between depression and occasion setting is warranted, and targeting this in the treatment of depression may be beneficial.

There are some limitations of our experiment. First, in Experiment 1, our measurement of self-report fear at transfer test likely suffered from extinction effects on those trials. Second, in Experiment 1, we counterbalanced the images for the CS+/CS-, auditory stimuli for occasion setters, and the training phase sequence (differential conditioning, feature-positive training, feature-negative training), which resulted in 24 counterbalanced versions of our experiment. While counterbalancing the CSs from occasion setting would have been preferable, we decided not to do this since this would have led to 48 counterbalanced versions with fewer than two participants per cell. Ultimately, participants learned reinforcement contingencies for all stimuli, suggesting that differences in the visual features of the stimuli did not produce any meaningful effects. Our counterbalancing in Experiment 2 improved on these limitations (see Methods). Third, our US expectancy rating prompt in Experiment 1 (rating US expectancy "in the next few moments") may have obscured direct occasion setter/US (no US) associations with this measure. This may have happened because the CS was temporally closer to the US than the occasion setter, and participants might have made their US expectancy rating during the CS, even though the occasion setter was a perfect predictor of the US.

Fourth, our design did not allow for the differentiation between delayed responding to the occasion setter vs responding to the CS after the occasion setter. To distinguish this, our experiment could have included feature-only trials to measure responding (e.g., SCR) at the occasion setter/US interval. This could have helped us further distinguish whether participants in

the feature-positive condition were exhibiting responses to the target after the feature or to the time interval after the feature (when the US was expected). However, much like our experiments, many previous occasion setting studies demonstrated occasion setting without the use of feature-only trials (e.g., Baeyens et al., 2001, 2004; Holland, 1984, 1986a). It is noteworthy that SCRs during the occasion-setter-trained CSs were significant and differential, and transfer test SCR was consistent with occasion setting, suggesting that the CSs acquired associative values with the US that were modulated by the putative occasion setters. Also, as expected, SCRs were stronger to the CSs than the occasion setters during compound trials when controlling for SCR on the first training block (see Supplementary Materials Figure SM4).

Fifth, there are many types of stimuli that are suitable for transfer tests, including trained/extinguished CS+s, CSs from separate occasion setting training, CSs from simultaneous "occasion setting" training, and others. As in our experiment, the CS+ (Baeyens et al., 2004; Holland, 1989b; Holland & Lamarre, 1984; Lamarre & Holland, 1985, 1987) and CS- (Holland, 1986a; Holland et al., 1999) have been used in multiple experiments as transfer test targets, and the CS+ or CS- have also been used as the only transfer test targets in some experiments (Holland et al., 1999; Lamarre & Holland, 1985). While we believe this approach is sufficient to demonstrate that occasion setting was learned, an even more definitive test would have included an additional set of separately trained positive and negative occasion setting stimuli (e.g., J→k+, k-; M→n-, n+) to investigate if occasion setters C and E produced successful transfer to conditional stimuli k and n, respectively. Lastly, the phrasing of the self-report questions in Experiment 1 may have facilitated participants' learning by asking participants to rate their fear of the CS alone or when preceded by the putative occasion setter. Future experiments could modify or omit this type of measurement.

In conclusion, our report is the first to investigate the effects of trait anxiety and trait depression on 1) differential conditioning, positive occasion setting, and negative occasion setting in 2) both fear conditioning and appetitive conditioning. In fear conditioning, we found that individuals high in trait anxiety had greater increases occasion setting fear relative to low trait anxiety. In appetitive conditioning, we found trait depression was associated with lower expectations of reward for the trained negative occasion setting compound and transfer of the negative occasion setter to the CS+. This report has many clinical implications, such as 1) exaggerated fear learning with occasion setting for anxious individuals, suggesting a potential need to target occasion setters in exposure therapy, and 2) under-expectation of reward in depressed individuals in relation to negative occasion setting, suggesting a potential need to enhance reward learning in treatment.

References

Acheson, D. T., Gresack, J. E., & Risbrough, V. B. (2012). Hippocampal dysfunction effects on context memory: Possible etiology for posttraumatic stress disorder. *Neuropharmacology*, *62*(2), 674–685. https://doi.org/10.1016/j.neuropharm.2011.04.029

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology*, *4*, 114–118. https://doi.org/10.1016/j.copsyc.2014.12.011

Amsterdam, J. D., Settle, R. G., Doty, R. L., Abelman, E., & Winokur, A. (1987). Taste and smell perception in depression. *Biological Psychiatry*, *22*(12), 1481–1485.

Antony, M. M., & Bieling, P. J. (1998). *Psychometric Properties of the 42-Item and 21-Item Versions of the Depression Anxiety Stress Scales in Clinical Groups and a Community Sample*. *10*(2), 176–181.

Baeyens, F., Vansteenwegen, D., Hermans, D., Vervliet, B., & Eelen, P. (2001). Sequential and simultaneous feature positive discriminations: Occasion setting and configural learning in human Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(3), 279–295. https://doi.org/10.1037/0097-7403.27.3.279

Baeyens, F., Vervliet, B., Vansteenwegen, D., Beckers, T., Hermans, D., & Eelen, P. (2004). Simultaneous and sequential Feature Negative discriminations: Elemental learning and occasion setting in human Pavlovian conditioning. *Learning and Motivation*, *35*(2), 136–166. https://doi.org/10.1016/S0023-9690(03)00058-4

Balea, P., Nelson, J. B., Ogallar, P. M., Lamoureux, J. A., Aranzubia-Olasolo, M., & Sanjuan, M. del C. (2020). Extinction contexts fail to transfer control: Implications for conditioned inhibition and occasion-setting accounts of renewal. *Journal of Experimental*

Psychology: Animal Learning and Cognition, *46*(4), 422–442. https://doi.org/10.1037/xan0000273

Barrett, J., & Armony, J. L. (2009). Influence of trait anxiety on brain activity during the acquisition and extinction of aversive conditioning. *Psychological Medicine*, *39*(2), 255–265. https://doi.org/10.1017/S0033291708003516

Berlin, I., Givry-Steiner, L., Lecrubier, Y., & Puech, A. J. (1998). Measures of anhedonia and hedonic responses to sucrose in depressive and schizophrenic patients in comparison with healthy subjects. *European Psychiatry*, *13*(6), 303–309. https://doi.org/10.1016/S0924-9338(98)80048-5

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology*, *9*(1), 65–73. https://doi.org/10.1016/j.coph.2008.12.014

Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State–Trait Anxiety Inventory, Trait version: Structure and content re-examined. *Behaviour Research and Therapy*, *36*(7–8), 777–788.

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. https://doi.org/10.1016/j.jbtep.2011.11.005

Bonardi, C., Robinson, J., & Jennings, D. (2017). Can existing associative principles explain occasion setting? Some old ideas and some new data. *Behavioural Processes*, *137*, 5–18. https://doi.org/10.1016/j.beproc.2016.07.007

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle Reflex Modification: Emotion or Attention? *Psychophysiology*, *27*(5), 513–522. https://doi.org/10.1111/j.1469-8986.1990.tb01966.x

Branco, L. D., Cotrena, C., Ponsoni, A., Salvador-Silva, R., Vasconcellos, S. J. L., & Fonseca, R. P. (2017). Identification and perceived intensity of facial expressions of emotion in bipolar disorder and major depression. *Archives of Clinical Neuropsychology*, *33*(4), 491–501.

Campbell-Sills, L., Norman, S. B., Craske, M. G., Sullivan, G., Lang, A. J., Chavira, D. A., Bystritsky, A., Sherbourne, C., Roy-Byrne, P., & Stein, M. B. (2009). Validation of a brief measure of anxiety-related severity and impairment: The Overall Anxiety Severity and Impairment Scale (OASIS). *Journal of Affective Disorders*, *112*(1), 92–101. https://doi.org/10.1016/j.jad.2008.03.014

Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., & Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *Journal of Anxiety Disorders*, *26*(3), 468–479. https://doi.org/10.1016/j.janxdis.2012.01.011

Castro, D. C., & Berridge, K. C. (2014). Advances in the neurobiological bases for food 'liking' versus 'wanting.' *Physiology & Behavior*, *136*, 22–30. https://doi.org/10.1016/j.physbeh.2014.05.022

Chambers, J. A., Power, K. G., & Durham, R. C. (2004). The relationship between trait vulnerability and anxiety and depressive diagnoses at long-term follow-up of Generalized Anxiety Disorder. *Journal of Anxiety Disorders*, *18*(5), 587–607. https://doi.org/10.1016/j.janxdis.2003.09.001

Chan, C. K. Y., & Lovibond, P. F. (1996). Expectancy bias in trait anxiety. *Journal of Abnormal Psychology*, *105*(4), 637–647. https://doi.org/10.1037/0021-843X.105.4.637

Clepce, M., Gossler, A., Reich, K., Kornhuber, J., & Thuerauf, N. (2010). The relation between depression, anhedonia and olfactory hedonic estimates—A pilot study in major depression. *Neuroscience Letters*, *471*(3), 139–143. https://doi.org/10.1016/j.neulet.2010.01.027

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., & Dour, H. J. (2016). Treatment for Anhedonia: A Neuroscience Driven Approach: 2015 ADAA Scientific Research Symposium: Treatment for Anhedonia. *Depression and Anxiety*, *33*(10), 927–938. https://doi.org/10.1002/da.22490

Craske, M. G., & Mystkowski, J. L. (2006). Exposure Therapy and Extinction: Clinical Studies. In *Fear and learning: From basic processes to clinical implications* (pp. 217–233). American Psychological Association. https://doi.org/10.1037/11474-011

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Craske, M. G., Wolitzky-Taylor, K. B., Mineka, S., Zinbarg, R., Waters, A. M., Vrshek-Schallhorn, S., Epstein, A., Naliboff, B., & Ornitz, E. (2012). Elevated responding to safe conditions as a specific risk factor for anxiety versus depressive disorders: Evidence from a longitudinal investigation. *Journal of Abnormal Psychology*, *121*(2), 315–324. https://doi.org/10.1037/a0025738

De Houwer, J., Crombez, G., & Baeyens, F. (2005). Avoidance Behavior Can Function as a Negative Occasion Setter. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 101–106. https://doi.org/10.1037/0097-7403.31.1.101

Declercq, M., & De Houwer, J. (2008). Evidence for the interchangeability of an avoidance behavior and a negative occasion setter. *Learning & Behavior*, *36*(4), 290–300. https://doi.org/10.3758/LB.36.4.290

Dibbets, P., Maes, J. H. R., & Vossen, J. M. H. (2002). Serial and simultaneous feature positive discriminations in a human Skinner box. *Behavioural Processes*, *59*(1), 1–14. https://doi.org/10.1016/S0376-6357(02)00046-3

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety*, *32*(4), 239–253. https://doi.org/10.1002/da.22353

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behavior Therapy*, *46*(5), 561–582. https://doi.org/10.1016/j.beth.2014.10.001

Franssen, M., Claes, N., Vervliet, B., Beckers, T., Hermans, D., & Baeyens, F. (2017). Reinstatement after human feature-positive discrimination learning. *Behavioural Processes*, *137*, 73–83. https://doi.org/10.1016/j.beproc.2017.01.007

Fraser, K. M., & Holland, P. C. (2019). Occasion setting. *Behavioral Neuroscience*, *133*(2), 145–175. https://doi.org/10.1037/bne0000306

Grillon, C., & Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology*, *38*(5), 807–815.

Haaker, J., Lonsdorf, T. B., Schümann, D., Menz, M., Brassen, S., Bunzeck, N., Gamer, M., & Kalisch, R. (2015). Deficient inhibitory processing in trait anxiety: Evidence from context-dependent fear learning, extinction recall and renewal. *Biological Psychology*, *111*, 65–72. https://doi.org/10.1016/j.biopsycho.2015.07.010

Holland, P. C. (1984). Differential effects of reinforcement of an inhibitory feature after serial and simultaneous feature negative discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*(4), 461–475. https://doi.org/10.1037/0097-7403.10.4.461

Holland, P. C. (1986a). Temporal determinants of occasion setting in feature-positive discriminations. *Animal Learning & Behavior*, *14*(2), 111–120. https://doi.org/10.3758/BF03200045

Holland, P. C. (1986b). Transfer after serial feature positive discrimination training. *Learning and Motivation*, *17*(3), 243–268. https://doi.org/10.1016/0023-9690(86)90013-5

Holland, P. C. (1989a). Acquisition and transfer of conditional discrimination performance. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(2), 154–165. https://doi.org/10.1037/0097-7403.15.2.154

Holland, P. C. (1989b). Transfer of negative occasion setting and conditioned inhibition across conditioned and unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(4), 311–328. https://doi.org/10.1037/0097-7403.15.4.311

Holland, P. C. (1989c). Feature extinction enhances transfer of occasion setting. *Animal Learning & Behavior*, *17*(3), 269–279. https://doi.org/10.3758/BF03209799

Holland, P. C. (1991a). Acquisition and transfer of occasion setting in operant feature positive and feature negative discriminations. *Learning and Motivation*, *22*(4), 366–387. https://doi.org/10.1016/0023-9690(91)90002-P

Holland, P. C. (1991b). Transfer of control in ambiguous discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(3), 231–248. https://doi.org/10.1037/0097-7403.17.3.231

Holland, P. C., & Lamarre, J. (1984). Transfer of inhibition after serial and simultaneous feature negative discrimination training. *Learning and Motivation*, *15*(3), 219–243. https://doi.org/10.1016/0023-9690(84)90020-1

Holland, P. C., Lamoureux, J. A., Han, J.-S., & Gallagher, M. (1999). Hippocampal lesions interfere with Pavlovian negative occasion setting. *Hippocampus*, *9*(2), 143–157. https://doi.org/10.1002/(SICI)1098-1063(1999)9:2<143::AID-HIPO6>3.0.CO;2-Z

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holmes, E. A., Lang, T. J., & Deeprose, C. (2009). Mental imagery and emotion in treatment across disorders: Using the example of depression. *Cognitive Behaviour Therapy*, *38*(S1), 21–28.

Jensen, D., Cohen, J. N., Mennin, D. S., Fresco, D. M., & Heimberg, R. G. (2016). Clarifying the unique associations among intolerance of uncertainty, anxiety, and depression. *Cognitive Behaviour Therapy*, *45*(6), 431–444. https://doi.org/10.1080/16506073.2016.1197308

Jovanovic, T., Kazama, A., Bachevalier, J., & Davis, M. (2012). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology*, *62*(2), 695–704. https://doi.org/10.1016/j.neuropharm.2011.02.023

Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J., & Phillips, M. L. (2005). The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry*, *58*(11), 843–853.

Kindt, M., & Soeter, M. (2014). Fear Inhibition in High Trait Anxiety. *PLOS ONE*, *9*(1), e86462. https://doi.org/10.1371/journal.pone.0086462

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., Brooks, N. H., & Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*, *43*(7), 1581–1588. https://doi.org/10.1038/s41386-018-0032-x

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain*, *131*(8), 2084–2093. https://doi.org/10.1093/brain/awn136

Laing, P. A. F., Vervliet, B., Fullana, M. A., Savage, H. S., Davey, C. G., Felmingham, K. L., & Harrison, B. J. (2021). Characterizing human safety learning via Pavlovian conditioned inhibition. *Behaviour Research and Therapy*, *137*, 103800. https://doi.org/10.1016/j.brat.2020.103800

Lamarre, J., & Holland, P. C. (1985). Acquisition and transfer of feature-negative discriminations. *Bulletin of the Psychonomic Society*, *23*(1), 71–74. https://doi.org/10.3758/BF03329783

Lamarre, J., & Holland, P. C. (1987). Transfer of inhibition after serial feature negative discrimination training. *Learning and Motivation*, *18*(4), 319–342. https://doi.org/10.1016/0023-9690(87)90001-4

Liberzon, I., & Sripada, C. S. (2007). The functional neuroanatomy of PTSD: A critical review. In E. R. De Kloet, M. S. Oitzl, & E. Vermetten (Eds.), *Progress in Brain Research* (Vol. 167, pp. 151–169). Elsevier. https://doi.org/10.1016/S0079-6123(07)67011-3

Lissek, S. (2012). Toward an Account of Clinical Anxiety Predicated on Basic, Neurally Mapped Mechanisms of Pavlovian Fear-Learning: The Case for Conditioned Overgeneralization. *Depression and Anxiety*, *29*(4), 257–263. https://doi.org/10.1002/da.21922

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*(5), 678–687. https://doi.org/10.1016/j.brat.2008.02.005

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized Anxiety Disorder Is Associated With Overgeneralization of Classically Conditioned Fear. *Biological Psychiatry*, *75*(11), 909–915. https://doi.org/10.1016/j.biopsych.2013.07.025

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391–1424. https://doi.org/10.1016/j.brat.2004.10.007

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*(1), 47–55. https://doi.org/10.1176/appi.ajp.2009.09030410

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, *14*(6), 417–428. https://doi.org/10.1038/nrn3492

Miller, R. R., & Oberling, P. (1998). Analogies between occasion setting and Pavlovian conditioning. In *Occasion setting: Associative learning and cognition in animals* (pp. 3–35). American Psychological Association. https://doi.org/10.1037/10298-001

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. *Trends in Cognitive Sciences*, *24*(3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review*, *32*(6), 496–509. https://doi.org/10.1016/j.cpr.2012.05.004

Mundy, E. A., Weber, M., Rauch, S. L., Killgore, W. D. S., Simon, N. M., Pollack, M. H., & Rosso, I. M. (2015). Adult Anxiety Disorders in Relation to Trait Anxiety and Perceived Stress in Childhood. *Psychological Reports*, *117*(2), 473–489. https://doi.org/10.2466/02.10.PR0.117c17z6

Norman, S. B., Hami Cissell, S., Means-Christensen, A. J., & Stein, M. B. (2006). Development and validation of an overall anxiety severity and impairment scale (OASIS). *Depression and Anxiety*, *23*(4), 245–249.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. *Annual Review of Psychology*, *68*(1), 73–100. https://doi.org/10.1146/annurev-psych-010416-044216

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Potts, A. J., Bennett, P. J., Kennedy, S. H., & Vaccarino, F. J. (1997). Depressive symptoms and alterations in sucrose taste perception: Cognitive bias or a true change in sensitivity? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *51*(1), 57.

Rescorla, R. A. (1986). Extinction of facilitation. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(1), 16.

Rømer Thomsen, K., Whybrow, P. C., & Kringelbach, M. L. (2015). Reconceptualizing anhedonia: Novel perspectives on balancing the pleasure networks in the human brain. *Frontiers in Behavioral Neuroscience*, *9*, 49.

Ruprecht, C. M., Wolf, J. E., Quintana, N. I., & Leising, K. J. (2014). Feature-positive discriminations during a spatial-search task with humans. *Learning & Behavior*, *42*(3), 215–230. https://doi.org/10.3758/s13420-014-0140-3

Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A., & Zarate Jr, C. A. (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *Journal of Psychiatric Research*, *44*(16), 1229–1235.

Scholten, W. D., Batelaan, N. M., Penninx, B. W. J. H., Balkom, A. J. L. M. van, Smit, J. H., Schoevers, R. A., & Oppen, P. van. (2016). Diagnostic instability of recurrence and the

impact on recurrence rates in depressive and anxiety disorders. *Journal of Affective Disorders*, *195*, 185–190. https://doi.org/10.1016/j.jad.2016.02.025

Scinska, A., Sienkiewicz-Jarosz, H., Kuran, W., Ryglewicz, D., Rogowski, A., Wrobel, E., Korkosz, A., Kukwa, A., Kostowski, W., & Bienkowski, P. (2004). Depressive symptoms and taste reactivity in humans. *Physiology & Behavior*, *82*(5), 899–904.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(4), 567–576.

Spielberger, C. D., & Gorsuch, R. L. (1983). *State-trait anxiety inventory for adults: Sampler set: manual, test, scoring key*. Mind Garden.

Staples-Bradley, L. K., Treanor, M., & Craske, M. G. (2018). Discrimination between safe and unsafe stimuli mediates the relationship between trait anxiety and return of fear. *Cognition and Emotion*, *32*(1), 167–173. https://doi.org/10.1080/02699931.2016.1265485

Starr, L. R., & Hershenberg, R. (2017). Depressive symptoms and the anticipation and experience of uplifting events in everyday life. *Journal of Clinical Psychology*, *73*(10), 1442–1461.

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology*, *39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

Trask, S., Thrailkill, E. A., & Bouton, M. E. (2017). Occasion setting, inhibition, and the contextual control of extinction in Pavlovian and instrumental (operant) learning. *Behavioural Processes*, *137*, 64–72. https://doi.org/10.1016/j.beproc.2016.10.003

Treadway, M. T., Bossaller, N. A., Shelton, R. C., & Zald, D. H. (2012). Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. *Journal of Abnormal Psychology*, *121*(3), 553.

van Dis, E. A. M., van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L. H., van den Heuvel, R. M., Cuijpers, P., & Engelhard, I. M. (2020). Long-term Outcomes of Cognitive Behavioral Therapy for Anxiety-Related Disorders: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *77*(3), 265–273. https://doi.org/10.1001/jamapsychiatry.2019.3986

van Vooren, P. R., Franssen, M., Beckers, T., Hermans, D., & Baeyens, F. (2012). Narrowing down the conditions for extinction of Pavlovian feature-positive discriminations in humans. *Learning & Behavior*, *40*(4), 393–404. https://doi.org/10.3758/s13420-011-0060-4

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., Schmidt, M., & Claes, S. (2013). Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry*, *73*(7), 639–645.

Wacker, J., Dillon, D. G., & Pizzagalli, D. A. (2009). The role of the nucleus accumbens and rostral anterior cingulate cortex in anhedonia: Integration of resting EEG, fMRI, and volumetric techniques. *Neuroimage*, *46*(1), 327–337.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form* [Other]. https://doi.org/10.17077/48vt-m4t2

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wong, A. H. K., & Lovibond, P. F. (2018). Excessive generalisation of conditioned fear in trait anxious individuals under ambiguity. *Behaviour Research and Therapy*, *107*, 53–63. https://doi.org/10.1016/j.brat.2018.05.012

Wu, H., Mata, J., Furman, D. J., Whitmer, A. J., Gotlib, I. H., & Thompson, R. J. (2017). Anticipatory and consummatory pleasure and displeasure in major depressive disorder: An experience sampling study. *Journal of Abnormal Psychology*, *126*(2), 149–159. https://doi.org/10.1037/abn0000244

Zinbarg, R. E., Mineka, S., Bobova, L., Craske, M. G., Vrshek-Schallhorn, S., Griffith, J. W., Wolitzky-Taylor, K., Waters, A. M., Sumner, J. A., & Anand, D. (2016). Testing a Hierarchical Model of Neuroticism and Its Cognitive Facets: Latent Structure and Prospective Prediction of First Onsets of Anxiety and Unipolar Mood Disorders During 3 Years in Late Adolescence. *Clinical Psychological Science*, *4*(5), 805–824. https://doi.org/10.1177/2167702615618162

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1911441 granted to Tomislav Zbozinek, PhD under the supervision of Dean Mobbs, PhD and Michael Fanselow, PhD. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.