Published in final edited form as:

Biol Psychol. 2019 January ; 140: 64–74. doi:10.1016/j.biopsycho.2018.11.009.

Sympathetic responding to unconditioned stimuli predicts subsequent threat expectancy, orienting, and visuocortical bias in human aversive Pavlovian conditioning

L. Forest Gruss^{a,b} and Andreas Keil^a

^aCenter for the Study of Emotion and Attention, Department of Psychology, University of Florida, Gainesville, Florida, USA

^bDepartment of Psychology, Vanderbilt University, Nashville, TN, USA

Abstract

Threat expectancy is the ability to predict an aversive outcome. What is not known is the influence of initial threat responding on the acquisition of verbal, attentional and perceptual biases towards conditioned threat cues. This study evaluated the extent to which initial unconditioned stimulus (UCS) responding was related to trial-by-trial self-reported expectancy, sensory processing (visuocortical EEG) and orienting (heart rate deceleration) to threat cues during extinction learning. Participants ($n=38$) viewed oriented Gabor gratings, associated with the presence (CS+) or absence (CS-) of a 96 dB white noise (UCS), flickering at 12 Hz to elicit steady state visually evoked potentials (ssVEPs). Multivariate multiple regression revealed greater initial UCS skin conductance responding to predict extinction responding: enhanced visuocortical discrimination, greater heart rate deceleration to CS+, and greater threat expectancy endorsements. These results suggest that the motivational intensity of initial threat reactivity (sympathetic UCS responding) drives learning-induced defensive dispositions across multiple response systems.

Keywords

Pavlovian conditioning; expectancy; threat bias; ssVEP; autonomic physiology

INTRODUCTION

Mammalian survival in threatening environments has necessitated the establishment of neurophysiological mechanisms adapted for evaluation of sensory events and subsequent action selection (Konorski, 1967). One of these mechanisms is the formation of associations between noxious outcomes and sensory events (threat cues). In the laboratory, this associative mechanism is examined in studies of classical (Pavlovian) aversive conditioning

Corresponding author: L. Forest Gruss, Vanderbilt University, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240, forest.gruss@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(e.g., Pavlov, 1960). These studies have converged to show that pairing a conditioned threat cue with a noxious outcome alters the organism's response to the threat cue over time at several levels of analysis (see Hamm & Weike, 2005, for a review), including autonomic measures (heart rate, skin conductance), sensory evoked brain potentials, and self-reported cognitive and affective evaluations (threat expectancy, hedonic valence ratings). One result of this literature is that there are pronounced inter-individual differences in acquiring conditioned responses, but the factors that contribute to these differences remain to be identified. Many theoretical and computational models of Pavlovian conditioning have aimed to explain the acquisition of conditioned responses as an interaction between (1) the motivational strength (or value) of the unconditioned stimulus (UCS), (2) the saliency of the conditioned stimulus, and (3) the adjustment of UCS expectancy to the CS- UCS contingencies (e.g., Rescorla & Wagner, 1972). The aim of the current study is to assess the extent to which the magnitude of an individual's response to a noxious event (the UCS) drives changes in the subsequent processing of a conditioned threat cue. Specifically, we hypothesize that reactivity to the UCS is an index of the motivational strength of the UCS and as such predicts responses to the conditioned stimulus. These responses are reflected as changes in sensory cortical processing, autonomic and reflexive engagement, and verbal expectancy, measured during the extinction phase of an uninstructed Pavlovian aversive conditioning paradigm.

Reactivity to the UCS as a predictor of conditioning magnitude across response systems.

The physical intensity of the UCS has been shown to moderate the strength of acquisition, robustness of the conditioned response and heightened resistance of acquired defensive tendencies during extinction learning (Prokasy et al., 1958; Walker, 1960; Dragoin, 1971; Rescorla & Wagner, 1972). In addition to the physical intensity of the UCS, recent research has begun to investigate the other side of the equation: To what extent the organism's reactivity to the UCS predicts conditioned responses. For example, increased stress and threat reactivity to the UCS have been suggested to result in enhanced association formation during information-integration (Ell et al., 2011) and enhanced memory consolidation during processing of words and unpleasant pictures (Weymar et al., 2012; Weymar et al., 2013). These findings give rise to specific hypotheses regarding temporal learning dynamics of conditioned threat cue processing: Greater responses to the UCS may index greater motivational value of the UCS and thus prompt more pronounced threat acquisition, measured as heightened autonomic, sensory, and self-report responses to the CS+ in extinction. Initial threat responding was measured in the current study in terms of defensive responding: skin conductance responding (sympathetic innervation) and heart rate acceleration (dually innervated by the autonomic nervous system) in response to the UCS. The following paragraph briefly introduces the various outcome measures along with their conceptual and empirical background.

Conceptual and empirical associations between visuocortical biases, expectancy, orienting and reflexive responding.

The formation of threat and safety associations has been shown to lead to learning-induced changes in sensory cortices (Headley & Weinberger, 2015; McTeague et al., 2015) and commensurate attention and perception biases towards threat cues that can be assessed at the

behavioral level (e.g., Lipp & Derakshan, 2005). The neurophysiological mechanisms of how sensory perception is modulated through experience or prediction from memory (Helmholtz, 1866) have recently been investigated (Summerfield & Egner, 2009; Egner et al., 2010; Browning & Harmer, 2012; Wyart et al., 2012; Clark, 2013; Blanco & Allen, 2018), specifically within the context of aversive conditioning (Cowen & Beier, 1950; Davey, 1992; Moratti & Keil, 2009; Kass et al., 2013). A core hypothesis, largely supported by this research, is that higher-order cortical regions such as orbitofrontal and anterior temporal cortices carry predictive information to sensory regions to optimize perception in a given context (Bar et al., 2006, McTeague et al., 2015).

Within this framework, the neural substrate of predictive signals may involve top-down bias signals modulating sensory and attention networks based on information derived from prior learning (Friston & Kiebel, 2009). In aversive conditioning, top-down modulation of sensory cortex may originate in anterior cortical as well as limbic structures such as the amygdaloid body, biasing sensory cortical processing based on prior experience (Skrandies & Jedynek, 2000; Amaral et al., 2003; Keil et al., 2009). These notions are consistent with classical views of aversive conditioning: For example, Epstein (1973) proposed expectancy to positively correlate with selective sensory responding. Visuocortical biases of heightened responding to the aversively conditioned stimulus (CS+) can be captured in primary visual cortex by means of the steady state visually evoked potential (ssVEP; Regan, 1989; Miskovic & Keil, 2012; Gruss et al., 2016). Manipulations of selective attention (Müller et al., 1998) and motivational relevance (Miskovic & Keil, 2012) prompt amplification of the ssVEP, which is reliably evoked by stimuli periodically and rapidly modulated in contrast or luminance. In aversive conditioning studies, the ssVEP is enhanced when evoked by a CS+ (aversively paired), compared to baseline or CS- (safety) stimuli. In the current study, our aim is to utilize the ssVEP as a measure for visuocortical engagement of the conditioned stimuli and to investigate to what extent initial threat reactivity modulates this evoked potential.

Prior work relating expectancy to sensory processing during threat learning has largely utilized self-reported expectancy ratings that reflect contingency awareness or has manipulated expectancy by systematically altering reinforcement probabilities. In a neurophysiological study, Moratti and Keil (2009) found enhanced electrocortical responses from visual cortex to reflect increased associative strength through CS-UCS pairings. Using fMRI, Kok and colleagues (2012) provided additional support for the notion that visual cortical activation relates to strength of association, reporting increased V1 activation with associative strength, but reduced activation with greater self-reported expectation. Here we aim to investigate the extent to which self-reported expectancy is predicted by the unconditioned response, instead of using expectancy as a predictor. We expect the unconditioned response to predict both CS-evoked ssVEP amplitude and self-reported expectancy.

Autonomic responses to conditioned threat cues may reflect attentional orienting as well as defensive engagement consistent with readiness for action as measured by defensive reflexes such as the eye blink startle response (Bradley et al., 2001). In the context of aversive conditioning, heart rate deceleration has been used as a measure of attention allocation, with

greater deceleration implicating enhanced attentional resources being committed to the threat-related cue through the influence of the parasympathetic branch of the autonomic nervous system (Bradley, 2009; Hodes et al., 1985). As a measure of sympathetic activation on the other hand, skin conductance responding is thought to reflect the degree to which the autonomic nervous system is preparing for action. Here, we expect these measures of orienting and defensive engagement in the extinction phase of conditioning to correlate with initial threat reactivity during acquisition.

The present study.

The present study implemented an uninstructed, discriminant classical conditioning paradigm with human observers. The discriminating feature for threat versus safety was the orientation of otherwise identical Gabor gratings, with a 96 dB white noise serving as the UCS. The main analysis focused on testing the central hypothesis, using multivariate multiple regression, that the initial response to the UCS predicts subsequent changes in expectancy (self-report), autonomic and defensive reflex engagement, and visuocortical bias formation (sVEPs), across individuals. In addition, we conducted cross-validation and complementary analyses on additional measures of defensive engagement across multiple levels of observations. To verify that conditioning effects were observed across response variables, univariate ANOVAs were conducted on each measure separately, as a method check. This approach enabled us to examine the multi-system impact of UCS responses on threat expectation across various components of defensive engagement, including sensory cortical, autonomic (sympathetic and parasympathetic), and verbal responses. Research into factors that influence aversive learning has often measured conditioning effects during the extinction phase of the experimental session (e.g., Myers & Davis, 2002). This approach is advantageous for physiological measurements due to the absence of UCS during that phase. It also informs regarding the persistence of learned responses, a question with clinical implications.

To evaluate the construct validity of our self-reported expectancy measure (Boddez et al., 2013) we used a canonical associative learning model that followed the Rescorla-Wagner (RW) Learning Rule (Rescorla & Wagner, 1972; Dayan & Abbott, 2001). The RW model incorporates changes in associative strength over time, in addition to a learning rate parameter that captures response changes due to contingency changes. In other words, unexpected or surprising changes in the reinforcement schedule are tracked using this model. By implementing a trial-by-trial learning rule, learning dynamics can be characterized by quantifying model parameters that best fit the empirical data of the expectancy ratings. Model parameters were then regressed against expectancy ratings in extinction to evaluate construct validity of the trial-by-trial self-reported expectancy measure used in the present study.

METHODS

Participants

Forty-nine undergraduates at the University of Florida participated in the current study for Psychology course credits. While aiming to recruit approximately 40 individuals based on

prior research investigating individual differences in autonomic measures and ssVEP data (McTeague et al., 2011; McTeague et al., 2017), some data were not included in the final analyses: One participant did not complete the study, three participants were not included due to technical difficulties during recording and 7 were excluded due to poor data quality in their electroencephalography (EEG) signal. The majority of the remaining 38 participants included in the analyses were right-handed (95%), male (55%), of average age 19.3 (1.2) and further demographic data were as followed: Caucasian (55%), Hispanic (29%), Asian (11%), and African-American (5%). All participants had normal or corrected vision and reported no personal or family history of photic seizures or epilepsy.

Stimuli

Visual stimuli were Gabor gratings (gray-scaled, sinusoidal gratings multiplied with a Gaussian envelope) of two orthogonal orientations ($\pm 5^\circ$ from the vertical meridian). The two orientations served as the discriminant CSs (conditioned stimuli) of the conditioning paradigm, only one orientation (the CS+) being paired with the UCS. Gabor patches were presented on a 23 inch LED monitor, 100 cm from the participant's head, subtending a vertical and horizontal visual angle of 10.9° and 14.3° , respectively, on a dark gray background, at a Michelson contrast of 0.98 and a spatial frequency of 0.6 cycles per degree. Gabor patches were flickered in a phase-reversal pattern at 12 Hz to drive the steady-state visually evoked potential (ssVEP) in visual cortex. One cycle of a Gabor patch and its phase-inverted counterpart therefore amounted to 166.66 ms to result in a 12 Hz oscillation. Forty-two cycles resulted in a trial length of 7000 ms throughout the experiment.

The aversive auditory stimulus serving as the unconditioned stimulus (UCS) in the current study was a 96 dB white noise filtered between 1 and 1000 Hz and presented from speakers placed 40 cm behind the participant's head for 1400 ms. Acoustic startle probes were presented from the same speakers for a length of 50 ms.

Design

The experiment consisted of a classical, differential conditioning paradigm with three phases: habituation, acquisition and extinction. No instructions were given regarding the contingency between the CSs and the UCS. During the habituation phase, participants were presented with the visual stimuli and were not exposed to the aversive stimulus (UCS). At the start of the acquisition phase, participants were told that occasionally they would hear loud noises and were asked after every trial to rate the likelihood of hearing that loud noise. Only the CS+ (right-slanted Gabor grating, $+5^\circ$) was paired with the UCS during the acquisition phase. The CS- was never paired with the UCS. Trial length during all three phases of the experiment was 7000 ms, unless the CS+ was paired with the UCS, in which case trials lasted an additional 1400 ms with the UCS occurring at 7000 ms and co-terminating with the CS+ at 8400 ms (delay conditioning). Extending the trial length to 8400 ms in CS+/UCS pairings therefore resulted in 50.5 Gabor grating cycles. Startle probes were presented throughout the experiment for two-thirds of the trials, 4000 or 5000 ms after CS onset, to elicit eye blink startle reflexes (see Figure 1).

At the start of the acquisition phase, a 100% reinforcement schedule was implemented, resulting in 7 CS+/UCS pairings (14 trials total). Subsequent pairings followed a partial reinforcement schedule, such that only 66% of the CS+ trials were accompanied by a UCS (again 7 CS+/UCS pairings, 22 trials total). Trials were pseudo-randomized to retain the same order of CS presentations across participants. The last phase of the experiment (extinction) was not made explicit to the participant; presentation of the CSs continued in the absence of the UCS. A re-exposure trial was introduced in the middle of extinction after 11 non-reinforced CS+ trials (22 trials total), resulting in one more CS+/UCS pairing. Forty-four trials in extinction, 36 trials in acquisition and 18 trials in habituation comprised the 98-trial uninstructed conditioning paradigm. Dependent measures were analyzed according to 5 experimental blocks: Habituation, Acquisition 1 (full reinforcement), Acquisition 2 (partial reinforcement), Extinction 1 (prior UCS re-exposure), and Extinction 2 (post UCS re-exposure). Utilizing changing reinforcement schedules, in addition to no contingency instruction to the participant, allowed for greater variability in behavioural responding reflecting contingency learning.

Procedure

Participants provided informed consent, were led to the laboratory and seated in the experiment room. After assessing net size for the electroencephalography (EEG), participants were given the State and Trait Anxiety Inventory (STAI) questionnaire (Spielberger et al., 1983) to fill out in the absence of the experimenter. Subsequently, Ag-AgCl electrodes for measuring skin conductance, heart rate and eye blink responses were attached in the following order, respectively: two 8 mm electrodes filled with 0.05 M NaCl paste on the hypothenar eminence of the left palm, 8 mm electrodes on each forearm, and two 4 mm electrodes under the left eye to record orbicularis oculi electromyogram (EMG). After fitting the EEG net and measuring impedances, lights were dimmed and brief instructions were given to the participant before commencing the experiment.

Following the habituation phase, participants rated the Gabor gratings using the Self Assessment Manikin (SAM) (Bradley and Lang, 1994) on perceived hedonic valence and arousal. Before continuing to the acquisition phase, participants were provided a mouse on a table to their right, where they could comfortably rest their arm. Participants were instructed to respond to a screen prompt that would follow each trial after 1500 – 3500 ms. The prompt asked the participant to rate the likelihood of hearing the loud noise when seeing the Gabor grating viewed on that same trial (expectancy rating) on a 10-point scale (1: unlikely, 5: uncertain, 10: likely). After providing their expectancy rating, the next trial would commence after 9000 – 11000 ms. At the conclusion of the experiment, the state anxiety portion of the STAI questionnaire was given to the participant to assess self-reported anxiety before and after aversive conditioning. Lastly, SAM ratings for the Gabor gratings were collected again, self-reported aversiveness of the UCS on a 10-point scale was recorded and participants were subsequently debriefed. The experiment lasted approximately 35 min. All procedures were approved by the institutional review board of the University of Florida and were in accordance with the Declaration of Helsinki.

Data processing

Electroencephalography (EEG)—EEG was recorded continuously using NetStation acquisition software, digitized at 500 Hz and band-pass filtered online from 0.05 to 48 Hz. Electrical Geodesic (EGI) HCGSN 129 channel sensor nets were used while keeping impedances below 40 k Ω . Offline data processing was implemented through EMEGS software (Peyk et al., 2011). As a first step, data was filtered with a 40 Hz lowpass filter (cut-off at 3 dB point, 45 dB/octave, 23rd order Butterworth) and a 3 Hz highpass filter (cut-off at 3 dB point, 18 dB/octave, 2nd order Butterworth) and segmented with 400 ms pre- and 7000 ms post-trigger for data epochs of 7400 ms length. Artifact rejection was subsequently executed by detecting artifacts in single channels in reference to Cz (recording reference) as deviations based on the distribution of the mean, the standard deviation, as well as the gradient of the voltage amplitude (Junghofer et al., 2000). Data were re-referenced to a global average and channels with artifacts were rejected and interpolated using a statistically weighted, spherical spline interpolation from the entire channel set. An average of 68 % of trials were retained after artifact rejection. Subsequently, trials were averaged together to represent condition-specific responses in the time domain.

To extract the steady state visually evoked potentials (ssVEPs) to the phase-reversed flickering Gabor patches a discrete Fourier transform (DFT) was performed on the condition averaged time-domain data. Two windows of the ssVEP were selected *a priori* for transformation into the frequency domain: an early window from 2–4 s and a late window from 5–7 s. The late window, immediately prior to UCS onset, is commonly utilized in the literature as conditioning effects are found during this time period (Moratti et al., 2006). The early window, on the other hand, provides an EEG signature prior to the acoustic startle probe (4 or 5 s post CS) and allows for investigation of temporal dynamics of ssVEP modulation within a trial, as well as over the experimental session. SsVEP responses were additionally averaged over an occipital sensor cluster encompassing Oz and its 6 nearest neighbors, including O1 and O2.

Autonomic and defensive reflex measures—Physiological measures were recorded via a computer running VPM software (Cook, 2002). Skin conductance was recorded continuously at a sampling rate of 20 Hz from two 8 mm Ag-AgCl electrodes filled with 0.05 M NaCl paste, placed on the left hypothenar eminence of the palm. Offline, skin conductance responding (SCR) was converted to microSiemens and averaged to half-second bins. To evaluate CS and UCS responding separately, two windows were *a priori* selected for subsequent data analysis. CS-related responding was extracted from maximum change scores from baseline (1 sec prior CS onset) between 2 and 7 sec post CS onset. UCS-related responding was extracted from maximum changes scores from baseline between 9 and 12 sec post CS onset. Both SCR values were then log transformed [$\log(\text{SCR}+1)$] and averaged over trials for condition-specific responses (CS+/-). Heart rate was recorded continuously from two 8 mm Ag-AgCl electrodes filled with electrolyte gel, placed on the inside of each forearm. Offline, phasic heart rate (HR) responses were assessed by reducing inter-beat (R-R) intervals to half-second bins (Graham, 1978), obtaining change scores from baseline (1 sec prior CS onset) and averaging trials for condition-specific responses (CS+/-). Peak heart rate acceleration (2–5 s) and deceleration (5–7 s) components of the triphasic waveform in

anticipation of the UCS were then extracted using difference scores: (1) initial deceleration (0–2 s) subtracted from peak acceleration, positive score, and (2) peak acceleration subtracted from peak deceleration prior to UCS, negative score (Hodes et al., 1985).

Eye blink startle magnitude to the acoustic startle probe was recorded via the left orbicularis oculi EMG from two 4 mm Ag-AgCl electrodes. Orbicularis oculi EMG activity was sampled at 1000 Hz, amplified, band-pass filtered (90–250 Hz) and integrated with a 20 ms time constant. Offline, startle magnitude was processed in VPM using a peak-scoring algorithm that utilizes onset latency and amplitude of each blink recorded (Balaban et al., 1986). Trials in which no blink was elicited were not included in subsequent analyses. Furthermore, 18 % of participants were qualified as nonresponders (average responding <9 %) and were excluded from further analyses (Blumenthal et al., 2005). Startle magnitudes were subsequently Z-scored due to individual variability, using mean and standard deviations within individuals across the entire experiment. Z-scores were trimmed by 3 standard deviations, converted to T scores and ultimately averaged across trials for condition-specific reflex responses (CS+/-). All condition-specific responses across peripheral physiological measures were averaged across the 5 experimental blocks described in the design subsection above.

Self-report measures—Self-reported expectancy ratings over 80 trials were converted from screen coordinate inputs to a 10-point scale representing the participant's ratings from unlikely to likely. To attain a quantitative measure of expectancy error, ratings were deviated from the contingency model of the experimental paradigm for the conditioned stimuli. For example, ratings after CS+ presentation in the acquisition phase should correspond to 10 (likely) according to the contingency parameters, whereas in the extinction phase ratings should change to 1 (unlikely). The contingency for CS- does not change throughout the experiment and should therefore be 1 (unlikely). As each individual will adjust their ratings throughout the experiment at different rates, these deviations from the contingency model reflected individual differences in expectancy dynamics on the level of verbal self-report. Raw scores from the 10-point scale were used for analysis of the induced alpha activity in extinction. By performing a median split on each individual's expectancy ratings, ratings above the median were classified as perceived 'threat'. Conversely, ratings below the median were classified as perceived 'safe' during the extinction phase.

The State and Trait Anxiety Inventory (STAI) questionnaire was completed by participants before the start of the experiment, as described in the procedure subsection above. Participants completed the State Anxiety portion of the questionnaire (Form Y-1) at the conclusion of the experiment to assess pre- to post-experimental changes in state anxiety. In addition, after the habituation phase and at the end of the experiment participants completed SAM ratings for the two CS stimuli to assess pre- to post-experimental changes in perceived hedonic valence and arousal. At the end of the experiment, UCS aversiveness was assessed on a 10-point scale.

Statistical procedures

For all analyses trials of each response measure were averaged into 5 experimental blocks as mentioned in the Design subsection: Habituation (H), Acquisition 1 (A1, full reinforcement), Acquisition 2 (A2, partial reinforcement), Extinction 1 (E1, pre UCS re-exposure), and Extinction 2 (E2, post UCS re-exposure).

Multivariate Multiple Regression—Under the hypothesis that threat reactivity to UCS exposure during initial aversive conditioning (acquisition) predicts subsequent modulations in outcome measures during extinction, a multivariate multiple regression was performed. This allowed us to account for 2 predictors in A1 – skin conductance reactivity and heart rate (HR) acceleration to the UCS – as well as 5 outcome measures related to CS+ responding in the first period of extinction: visuocortical responding (EEG), HR deceleration, eye blink startle magnitude, skin conductance response (SCR), and expectancy ratings. All five dependent variables were selected to address components related to expectancy, selective perceptual, attentive and reflexive processing. State and trait anxiety measures from the STAI questionnaire were additionally evaluated in a separate univariate regression analysis.

Univariate Analyses—Univariate analyses were performed as a method check to track conditioning effects across time: Rather than testing hypotheses, these analyses aimed to establish the presence of typical, well-known, effects of aversive conditioning on each dependent measure, before conducting regression analyses, which were the main focus of this study. Univariate repeated measures ANOVAs were performed on each dependent measure with the within-subjects factors of *CS condition* (2, CS+/-) and *block* (4, A1, A2, E1, E2). Follow-up repeated measures ANOVAs were then performed with the factors *CS condition* and a two-level factor *block* to quantify conditioning dynamics during the acquisition phase (A1/A2), the extinction phase (E1/E2), as well as the transition into extinction (A2/E1). The same statistical procedure was performed on the behavioural data (expectancy ratings) to verify CS discrimination on the group level, with the exception of the State Anxiety portion of the STAI-Q that was analyzed with a paired samples t-test (before and after experiment). SAM ratings of the CSs were analyzed with repeated measures ANOVA for both Valence and Arousal, with the within-subjects factors of *CS condition* (2, CS+/-) and *time* (2, before and after experiment). Results of these analyses are reported below, to allow comparison with a large body of previous studies, and to enable readers to assess the robustness of each measure's sensitivity to the present conditioning protocol.

Correspondence of ratings with the Rescorla-Wagner model—For the Rescorla-Wagner (RW) model we formulated a weight vector that is updated every trial as followed: $w(t+1) = w(t) + \eta * \delta * U(t)$, where η is a learning rate parameter (fixed), δ the prediction error (changing), U a stimulus saliency vector (fixed) and t the trial count. The prediction error, δ , changes with each trial as the goal of the weight vector is to bring the prediction error to zero, as $\delta = r(t) - w(t) * u(t)$, r being the actual outcome and the $w(t) * u(t)$ term being the expected outcome. When the actual outcome meets expectation, all conditioned stimuli and their outcomes are fully predicted. In the present case, r will be modeled via the pseudo-randomized reinforcement schedule, i.e. the presence or absence of the UCS on each

trial (binary vector). Likewise, the U parameter is determined in part by the trial order of the experiment: U represents stimulus saliency, i.e. saliency of the CS+ and CS- Gabor patches. As the ability to perceptually discriminate these two stimuli may vary by individual (right versus left orientation) a two-vector (CS+/-) parameter was implemented in which the U parameter represented a similarity index. For example, the similarity of CS+ on trial (t) to CS- generally may be represented as $[1 \ 0.3]$, respectively. The greater the difference between these two values (0.7 in this case), the more perceptually distinct they are, and vice versa. The η parameter, on the other hand, represents a learning parameter or associability of the stimulus with the reinforcement (UCS), which again will differ by individual.

The U and η parameters were then adjusted for each participant to yield the highest correlation with their observed expectancy ratings. Correlating the model weight vector against the empirical data provides a correlation coefficient. Executing the Rescorla-Wagner Learning Rule across a range of U and η parameters (50 each, 2,500 models per person) then allowed us to attain the RW model parameters best suited for each participant, yielding the highest correlation coefficient. These parameters were then regressed against the self-reported expectancy measure in extinction.

RESULTS

Multivariate Multiple Regression

A multivariate multiple regression was performed with 2 independent variables (skin conductance responding (SCR) and heart rate (HR) acceleration to the UCS) predicting 5 dependent variables (visuocortical responding (EEG), HR deceleration, eye blink startle magnitude, skin conductance responding (SCR), and verbal expectancy ratings). A temporal component was included in the analysis with independent variables being taken from the initial acquisition phase (A1) and dependent variables from the first extinction phase (E1). This statistical procedure specifically tested the extent to which threat reactive measures during initial threat exposure successfully predict response outcomes later in extinction.

Pillai's trace for the omnibus regression with 2 independent and 5 dependent variables was .82 [$F(10,38) = 2.61$, $p < .05$], suggesting the two UCR reactivity measures combined explained a significant portion of the variance seen in the outcome measures. Of the two predictors, however, only SCR to the UCS was a significant predictor [$F(5,18) = 4.03$, $p < .05$], with the multivariate test for the HR acceleration predictor not being significant [$F(5,18) = 1.54$, $p = .23$]. Omnibus univariate tests for each of the dependent variables revealed a significant R^2 for HR deceleration in extinction learning [$R^2 = .47$, $F(2,24) = 9.70$, $p < .01$] and the expectancy ratings [$R^2 = .29$, $F(2,24) = 4.42$, $p < .05$], with ssVEP not being jointly predicted by both independent variables [$R^2 = .19$, $F(2,24) = 2.58$, $p = .10$]. However, Because the UCS-SCR measure was shown to be the significant predictor among independent variables, with no effects observed for HR acceleration to the UCS, we followed up with a multivariate test of SCR alone predicting the outcome measures: SsVEP discrimination in extinction learning was significantly predicted by initial SCR reactivity [$F(1,22) = 5.16$, $p < .05$], as was HR deceleration [$F(1,22) = 18.92$, $p < .001$]. Startle and CS-SCR were not predicted as significant outcome measures [$R^2 = .17$, $F(2,24) = 2.27$, $p = .13$; $R^2 = .08$, $F(2,24) = .96$, $p = .40$]. These results suggest that SCR to the UCS during

initial acquisition plays a significant role in predicting ssVEP discrimination and HR deceleration during extinction learning. Figure 2 illustrates SCR magnitude in A1, splitting responders into tertiary groups ($n=11$), scatterplots of the related outcome measures in E1, as well as scalp topographies of the ssVEP enhancement to CS+ compared to CS-. Participants with greater skin responses to the UCS during A1 demonstrated subsequent enhancement in their visual engagement to the CS+ over the CS- stimulus at the onset of extinction (Figure 2c).

Participants with greater unconditioned responding to the UCS at the onset of conditioning endorsed, subsequently, greater expectancy errors of the UCS in extinction at the presentation of the CS+ (Figure 3, illustrating high and low reactivity groups). Greater SCR magnitude to the UCS predicted greater HR deceleration to the CS+ in extinction (Figure 4).

An additional multivariate multiple regression was performed with the previous 2 independent variables and 3 dependent variables from the questionnaire data (State Anxiety ratings pre and post experiment, as well as Trait Anxiety ratings). Pillai's trace (.34) for the overall multivariate test was not significant [$F(6,56) = 1.88, p = .10$], nor were multivariate tests for each predictor [SCR: $F(3,27) = 2.66, p = .07$; HR acceleration: $F(3,27) = .89, p = .46$].

Univariate analyses

Visuocortical responding—For the early ssVEP window (2–4 s), repeated measures ANOVA revealed a main effect of the 4-level factor *block* [$F(3,111) = 5.10, p < .01$], such that greater ssVEP enhancement was seen earlier on during the experiment than towards the end [linear contrast: $F(1,37) = 8.61, p < .01$]. The *CS condition* main effect was not significant across the four experimental blocks. When breaking down into a 2-level factor of *block*, however, there was a main effect of *block* [$F(1,37) = 9.76, p < .01$] as before, with A1 demonstrating greater ssVEP enhancement than A2, as well as an interaction with *CS condition* [$F(1,37) = 4.85, p < .05$]. Here, the ssVEP response to CS- significantly decreased from A1 to A2 [$t(37) = 4.25, p < .001$]. In addition, a main effect of *CS condition* was found in the transition period from A2 to E1 [$F(1,37) = 5.81, p < .05$], such that the ssVEP response to CS+ was significantly enhanced.

For the late ssVEP window (5–7 s), repeated measures ANOVA demonstrated no significant main effects over the 4-level factor of *block*. For the acquisition blocks, however, an interaction with the 2-level factor *block* and *CS condition* [$F(1,37) = 4.27, p < .05$] revealed the ssVEP response to CS+ during A1 to be significantly enhanced over the CS- response [$t(37) = 2.25, p < .05$] (Figure 5; means in Table 1).

Autonomic and Reflex Physiology—Repeated measures ANOVA showed that eye blink startle magnitude significantly decreased over the course of the experiment, main effect of *block* [$F(3,69) = 44.03, p < .001$; linear contrast: $F(1,23) = 65.84, p < .001$]. A main effect of *CS condition* was also observed [$F(1,23) = 17.88, p < .001$] such that greater startle magnitude for CS+ trials over CS- were evident. When considering specific blocks, the same main effects of *block* [$F(1,28) = 56.27, p < .001$] and *CS condition* [$F(1,28) = 7.52, p < .05$] were significant for the acquisition blocks, as well as over the transition period from A2 to

E1, *block* [$F(1,27) = 9.78, p < .01$], *CS condition* [$F(1,27) = 17.95, p < .001$]. In the extinction phase, main effects of *block* [$F(1,25) = .48, n.s.$] and *CS condition* [$F(1,25) = 2.05, n.s.$] were no longer significant (Figure 6).

Repeated measures ANOVA on heart rate (HR) deceleration showed a main effect of *block* [$F(3,108) = 4.01, p < .05$] by which greater deceleration was found during the A2 and E1 blocks [quadratic contrast: $F(1,36) = 10.24, p < .01$]. A main effect of the 2-level factor *block* [$F(1,36) = 9.43, p < .05$] during the acquisition phase further demonstrated the increased HR deceleration in A2 over A1. However, main effect of *CS condition* was not significant [$F(1,36) = 4.06, p = .051$]. Over the A2 and E1 blocks, a main effect of *CS condition* did become apparent [$F(1,36) = 4.35, p < .05$] (Table 1). No significant effects were found for HR acceleration.

Skin conductance responses (SCR) analyzed with a 4-level factor *block* did not reveal any main effects in the repeated measures ANOVA. However, a main effect with the two-level factor *block* was found during acquisition [$F(1,33) = 4.77, p < .05$], in which SCR decreased over the acquisition phase. During the transition from A2 to E1 a main effect of *CS condition* was revealed [$F(1,33) = 5.27, p < .05$], indicating enhanced SCR to the CS+. In addition, a main effect of *block* was found [$F(1,33) = 4.87, p < .05$], but in the opposite direction as was found during acquisition. Here, at the start of the extinction phase anticipatory SCR increases again. No main effects or interactions were reported for the two blocks in the extinction phase. Lastly, SCR to the UCS was analyzed and a main effect of the 4-level factor *block* was found [$F(3,96) = 16.11, p < .001$] as well as the main effect of *CS condition* [$F(1,32) = 38.38, p < .001$], by which SCR was enhanced for CS+ trials overall and was more pronounced at the beginning of the experiment. In addition, an interaction of *block* by *CS condition* was found [$F(3,96) = 15.42, p < .001$], in which responses to CS+ trials decreased over blocks, but not for responses to CS-trials (Table 1).

Self-report measures—To assess CS discrimination on the group level, across all 4 block of the experiment where expectancy ratings were given, a repeated measures ANOVA was implemented. A main effect of the 4-level factor *block* [$F(3,111) = 43.02, p < .001$], of *CS condition* [$F(1,37) = 102.08, p < .001$] and an interaction of *block* by *CS condition* was found [$F(3,111) = 11.56, p < .001$], demonstrating not only clear endorsement of greater expectancy for CS+ trials (*CS condition* main effect), but that ratings for CS+ decreased over the course of the experiment ('less likely') compared to ratings for CS- trials which remained stable.

The State Anxiety portion of the STAI was analyzed with a paired samples t-test, and participants endorsed significantly higher state anxiety after than before the experiment [$t(37) = 3.63, p < .01$]. The repeated measures ANOVA on the SAM ratings revealed that participants rated CS+ (main effect of *CS condition*) as significantly less pleasant [$F(1,37) = 31.01, p < .001$] and more arousing than the CS- [$F(1,37) = 12.59, p < .01$]. However, *CS condition* interacted with *time* for both Valence and Arousal [$F(1,37) = 21.82, p < .001$; $F(1,37) = 17.24, p < .001$, respectively] in which decreased pleasantness and increased arousal for the CS+ compared to CS- was greater after the experiment [$t(37) = -5.56, p < .001$; $t(37) = 4.62, p < .001$, respectively]. In addition, a main effect of *time* was found for

Arousal [$F(1,37) = 10.98, p < .01$] such that greater arousal endorsements occurred after the experiment, corroborating findings of increased state anxiety after the experiment. Ratings of the UCS averaged to 7.2 (2.2) on a 10-point scale of increasing aversiveness.

Correspondence of ratings with prediction from the Rescorla-Wagner model

Two regression analyses were performed with the RW model parameters U and η predicting the expectancy error as a validity check. Correspondence with the prediction of the RW model would support that the present ratings tap into constructs proposed by this model, e.g. learning rate, η , and saliency U . The U parameter was not found to be significantly related to the expectancy measure in E1 [$F(1,37) = .26, n.s.; R^2 = .01$]. The η parameter, however, was significantly related to the expectancy measure [$F(1,37) = 9.34, p < .01; R^2 = .21, \beta = -.45$] such that smaller η values predicted larger expectancy errors. As smaller learning rates indicate less adjustment to contingency changes during the experiment, this corresponded to greater expectancy errors in the CS+ during extinction, as participants continued to rate the UCS as 'likely to happen' despite the new reinforcement schedule (extinction). On the other hand, larger learning rates indicate greater adjustment to contingency changes, which leads the prediction error to converge quicker to zero (greater predictability). Larger learning rates therefore did significantly predict smaller expectancy errors in extinction trials to CS+. Figure 7 illustrates an example of our RW model as a weight vector over time with the fixed r vector, a fixed $U(0.5)$ and two different η values, as well as the relationship between the η parameter and the expectancy error of the CS+.

DISCUSSION

The aim of the current study was to evaluate the impact of UCS responding (UCR) during threat acquisition on sensory, attentive and reflexive processing of threat cues and self-reported expectancy in extinction. Greater UCR has been traditionally associated with enhanced aversive learning, reflected in accompanying verbal threat expectancy reports (Rescorla & Wagner, 1972; Epstein, 1973). Here, we examined the extent to which the motivational strength of the UCS, established during initial threat acquisition as UCR, drives multisystem threat processing during extinction learning. Specifically, we tested the hypothesis that verbal expectancy, visuocortical, autonomic and reflexive indices of threat processing are predicted by the strength of the autonomic response to the UCS. A multivariate multiple regression model showed that the sympathetic (specifically skin conductance) response to the UCS during initial conditioning predicted selective CS+ processing during extinction: Greater sympathetic responding to the UCS predicted enhanced visuocortical threat discrimination, greater HR deceleration, and higher endorsements of verbal expectancy during extinction learning. By contrast, heart rate (HR) acceleration to the UCS did not predict variance in any of the outcome measures. These results suggest that threat biases in visual cortex, heart rate orienting, and persistent self-reports of UCS expectancy may be driven by inter-individual differences in initial defensive reactivity.

Reliable conditioning effects were found across all neurophysiological and behavioural measures, serving as a method check and validating that the present experimental procedure

resulted in aversive conditioning. However, temporal dynamics varied across measures, such that discrimination between CS+ and CS- occurred after few trials in some measures (startle response, late window ssVEP and expectancy ratings), and appeared only in the second half of acquisition in others (heart rate deceleration, SCR and early window ssVEP). Measures showing discrimination in the second part of acquisition further persisted into the first extinction block. Interestingly, these same measures were predicted in extinction to vary by sympathetic UCR magnitude during initial conditioning. These temporal dynamics will be explored next.

Evolution of conditioning effects across the experimental session.

The visuocortical signature of the conditioned stimuli (CS), the ssVEP, displayed rapid initial conditioning effects in the first half of acquisition for the late recording window (5–7 s). Visuocortical enhancement to the CS+ in this window has been interpreted as reentrant, top-down modulation from frontal-parietal cortical networks, characterized as motivated attention to the impending UCS (Moratti et al., 2006; Petro et al., 2017). SsVEP discrimination later in acquisition subsequently shifted to an earlier window during the trial, which may be related to the frequent occurrence of startle probes mid-trial. These probes suppressed visuocortical activity as shown in Figure 5. In a conditioning task with ssVEPs, Miskovic and Keil (2013) argue that the associative strengthening of CS-UCS associations in primary visual cortex is facilitated with temporal proximity between CS and UCS. The startle probes share several characteristics with the UCS, including physical structure (white noise), loudness and, potentially, aversiveness. Thus, their mid-trial presence may have prompted ssVEP discrimination to occur earlier in the trial than previously observed in studies with no startle probes. As learning progressed, threat evaluation through visual discrimination occurred as soon as the CSs were presented. Future studies may further investigate this temporal shift in the ssVEP trial window as aversive learning progresses, in paradigms that do not involve acoustic startle probes.

Sympathetic response magnitude predicts subsequent orienting, sensory processing, and expectancy.

The current results demonstrate that initial sympathetic response magnitude (SCR to the UCS) predicts outcome measures in the subsequent extinction phase. HR acceleration, however, did not predict outcome measures in the current model. An explanation for this dichotomy in reactivity measures may be in part due to the innervation of the autonomic nervous system (ANS): The heart is dually innervated by the sympathetic and parasympathetic branches of the ANS, whereas skin conductance reactivity is solely mediated by the sympathetic ANS. This highlights the important role of sympathetic reactivity for associative learning. Future work may explore this differentiation within the ANS and the role of sympathetic arousal for driving changes in verbal expectancy, orienting, and visuocortical biases during different conditioning tasks, and in different populations.

Eye blink startle magnitude was not predicted by threat reactivity. In previous work, startle modulation (Hamm & Vaitl, 1996) and eyeblink conditioned responses (Perruchet, 1985) were more closely related to associative strength throughout the learning period irrespective of reported contingency awareness. Therefore, the current study would suggest expectancy

to not merely be an establishment of associative strength, but instead appears more focused on heightened threat orienting through perceptual processing and attentional orienting (HR deceleration). The lack of significant findings for the anticipatory SCR response in extinction may be due to weak signal strength. Stimulus evoked SCR responding to the UCS is markedly stronger than the anticipatory period prior to UCS onset (Figure 2a).

Prior dispositions to threat, such as clinical diagnoses, lend further credence to the formation of threat biases early on during threat exposure and the subsequent strengthening of conditioned responses and failure to extinguish these responses. In a series of experiments, Davey and Dixon (1996) illustrate that individuals high in fear not only demonstrate an *a priori* bias towards threat, but also an *a posteriori* bias (overestimating threat contingencies already experienced). Work by Wiedemann and colleagues (2001) further corroborated these findings with panic patients who displayed an overestimation of threat expectancies and contingent consequences. In a conditioning paradigm using a blocked CS stimulus, Boddez and colleagues (2012) found individuals high in trait anxiety to show deficient blocking, suggestive of a heightened threat bias and accompanying threat generalization. While these studies support the notion that prior dispositions of threat sensitivity lead to behavioural threat biases, work in the affective neurosciences illustrate enhanced visuocortical activation (ssVEP and ERP components) towards perceived threat stimuli as well (McTeague et al., 2011; Wieser et al., 2011; Deweese et al., 2014; Weymar et al., 2014). Unique to the present study, however, was utilizing threat reactivity in a group of non-anxious individuals to investigate their threat biases and expectations on extinction learning.

A clear limitation of the current study lies in the inferences that can be drawn from expectancy ratings. Work in experimental psychology in humans has demonstrated the ambiguity of verbal behaviour in relation to neurophysiological measures in emotion and cognitive research (Lacey, 1959; Lang, 1968). To further explore the relationship between neurophysiological mechanisms and verbal behaviour in the context of threat expectation, more systematic experimentation with verbal prompts would be necessary. The value of subjective expectancy ratings has been questioned in the literature for its validity in aversive conditioning work (Boddez et al., 2013). To assess construct validity of our expectancy measure in the current study, we implemented regression analyses of model parameters that follow the Rescorla-Wagner Learning Rule (Dayan & Abbott, 2001) against expectancy errors in extinction. Model parameters (U and η) were individualized based on the empirical data. The learning rate parameter (η) was significantly related to expectancy errors in extinction, such that a smaller learning rate corresponded to greater errors, i.e. extinction resilience.

Although the present analysis of rating validity was supported by the Rescorla and Wagner model, neurophysiological measures may be more suitably modeled by other reinforcement learning algorithms (Sutton & Barto, 1998; Dayan & Abbot, 2001). The study of threat expectancy during aversive learning may therefore require a more integrative, multi-systems approach in understanding how threat expectancies are acquired, retained and modified through learning. The temporal evolution of associative learning may differ across measures and may be differentially impacted by other factors such as attention and motivation. Startle responding for example does not appear to be impacted by contingency awareness (Hamm &

Vaitl, 1996), and may therefore be better suited for indexing physiological aspects of aversive conditioning in a multi-system model of aversive conditioning. Work comparing computational models may therefore benefit from exploring multimodal dynamics of threat expectancy. In particular, extended extinction may be explored in next-day threat recall or threat re-exposure for learning dynamics across response systems (Panitz et al., 2015).

The present findings illustrate the relevance of threat acquisition for responses during subsequent extinction learning: Individuals who fail to adapt to changing contingencies display heightened threat responding in extinction. These findings suggest that verbal expectancy ratings carry potentially valuable information as one component of threat expectation. The clinical relevance for a more integrative framework in aversive learning, including temporal learning dynamics, threat reactivity and verbal expectancy ratings is exemplified in NIMH's RDoC (Research Domain Criteria) Initiative (e.g., Insel & Cuthbert, 2009). By utilizing a dimensional and multi-systems approach, a greater understanding between neurophysiological mechanisms and behavioural symptomatology can be attained for fear and anxiety disorders (Lissek et al., 2005; Britton et al., 2011; McTeague & Lang, 2012; Lissek et al., 2014). In the current study, individual differences in initial threat reactivity, measured as sympathetic responding, have a significant impact on threat evaluation (expectancy) in extinction. Future work investigating individual dispositions (clinical populations, genotype differences) may shed light on these learning dynamics during aversive conditioning.

Defensive reactivity includes responses of the entire organism that dispose mammals to effectively meet environmental challenges and threats and may best be appreciated as an evolutionarily driven motivator for action preparation (Lang et al., 1998). Heightened expectancy and the associated predictive top-down biasing of attention/perception systems may therefore be understood as an evolutionarily driven motivator for rapid, accurate threat detection (Epstein, 1973). Future studies may wish to explore the extent to which visuocortical signals retain an expectancy engram after repeated threat exposure or alternatively are merely modulated by top-down hierarchical signals from anterior cortical regions. The current findings support the notion that threat processing involves an ensemble of neurophysiological processes preparing the animal for adaptive action under threat. Moreover, sympathetic responding during initial threat exposure significantly shapes threat processing during later uncertain threat conditions. Together, these findings contribute to our understanding of the role of the unconditioned response magnitude in aversive conditioning, vis-à-vis the different levels affected by fear conditioning, including self-reported expectancy, and physiological adaptations (Pavlov, 1960; Rescorla & Wagner, 1972; Ohman & Mineka, 2001).

ACKNOWLEDGMENTS

This work was supported by the National Institute of Health [MHO97320]. The authors declare no competing interests.

REFERENCES

- Amaral DG, Behniea H, & Kelly JL (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience*, 118(4), 1099–1120. [PubMed: 12732254]
- Balaban M, Losito B, Simons R, & Graham F (1986). Off-line latency and amplitude scoring of the human reflex eye blink with Fortran IV. *Psychophysiology*, 23(5), 612.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, ... Halgren E (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*, 103(2), 449–454. doi:10.1073/pnas.0507062103 [PubMed: 16407167]
- Bechara A, Tranel D, Damasio H, Adolphs R, Rockland C, & Damasio AR (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, 269(5227), 1115–1118. [PubMed: 7652558]
- Blanco NDC, & Allen JJB (2018). Guidance of spatial attention during associative learning: contributions of predictability and intention to learn. *Psychophysiology*, e13077. [PubMed: 29572912]
- Blumenthal TD, Cuthbert BN, Filion DL, Hackley S, Lipp OV, & van Boxtel A (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, 42(1), 1–15. doi:10.1111/j.1469-8986.2005.00271.x [PubMed: 15720576]
- Boddez Y, Baeyens F, Luyten L, Vansteenwegen D, Hermans D, & Beckers T (2013). Rating data are underrated: validity of US expectancy in human fear conditioning. *J Behav Ther Exp Psychiatry*, 44(2), 201–206. doi:10.1016/j.jbtep.2012.08.003 [PubMed: 23207968]
- 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. *J Behav Ther Exp Psychiatry*, 43(2), 832–837. doi:10.1016/j.jbtep.2011.11.005 [PubMed: 22197754]
- Bolles RC, & Fanselow MS (1980). A perceptual-defensive-recuperative model of fear and pain. *Behavioral and Brain Sciences*, 3, 291–323.
- Bradley MM (2009). Natural selective attention: orienting and emotion. *Psychophysiology*, 46(1), 1–11. doi:10.1111/j.1469-8986.2008.00702.x [PubMed: 18778317]
- Bradley MM, Codispoti M, Cuthbert BN, & Lang PJ (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276–298. [PubMed: 12934687]
- Bradley MM, & Lang PJ (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry*, 25(1), 49–59. [PubMed: 7962581]
- Britton JC, Lissek S, Grillon C, Norcross MA, & Pine DS (2011). Development of anxiety: the role of threat appraisal and fear learning. *Depress Anxiety*, 28(1), 5–17. doi:10.1002/da.20733 [PubMed: 20734364]
- Browning M, & Harmer CJ (2012). Expectancy and surprise predict neural and behavioral measures of attention to threatening stimuli. *Neuroimage*, 59(2), 1942–1948. doi:10.1016/j.neuroimage.2011.09.007 [PubMed: 21945791]
- Büchel C, Morris J, Dolan RJ, & Friston KJ (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron*, 20(5), 947–957. [PubMed: 9620699]
- Clark A (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*, 36(3), 181–204. doi:10.1017/S0140525X12000477 [PubMed: 23663408]
- Cook EW, 3rd. (2002). VPM reference manual. Birmingham, AL: Author. Cowen, E. L., & Beier, E. G. (1950). The influence of “threat-expectancy” on perception. *J Pers*, 19(1), 85–94.
- Davey GC, & Dixon AL (1996). The expectancy bias model of selective associations: the relationship of judgements of CS dangerousness, CS-UCS similarity and prior fear to a priori and a posteriori covariation assessments. *Behav Res Ther*, 34(3), 235–252. [PubMed: 8881093]
- Davey GCL (1992). An expectancy model of laboratory preparedness effects. *Journal of Experimental Psychology: General*, 121(1), 24–40.
- Davis M (1992). The role of the amygdala in fear and anxiety. *Annu Rev Neurosci*, 15, 353–375. doi:10.1146/annurev.ne.15.030192.002033 [PubMed: 1575447]

- Dayan P, & Abbott LF (2001). Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA, USA: MIT Press.
- Deweese MM, Bradley MM, Lang PJ, Andersen SK, Müller MM, & Keil A (2014). Snake fearfulness is associated with sustained competitive biases to visual snake features: hypervigilance without avoidance. *Psychiatry Res*, 219(2), 329–335. doi:10.1016/j.psychres.2014.05.042 [PubMed: 24930577]
- Dragoin WB (1971). Conditioning and extinction of taste aversions with variations in intensity of the CS and UCS in two strains of rats. *Psychonomic Science*, 22(5), 303–305.
- Egner T, Monti JM, & Summerfield C (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci*, 30(49), 16601–16608. doi:10.1523/JNEUROSCI.2770-10.2010 [PubMed: 21147999]
- Ell SW, Cosley B, & McCoy SK (2011). When bad stress goes good: increased threat reactivity predicts improved category learning performance. *Psychon Bull Rev*, 18(1), 96–102. doi:10.3758/s13423-010-0018-0 [PubMed: 21327364]
- Epstein S (1973). Expectancy and magnitude of reaction to a noxious UCS. *Psychophysiology*, 10(1), 100–107. [PubMed: 4684224]
- Fanselow MS, & Poulos AM (2005). The neuroscience of mammalian associative learning. *Annu Rev Psychol*, 56, 207–234. doi:10.1146/annurev.psych.56.091103.070213 [PubMed: 15709934]
- Friston K, & Kiebel S (2009). Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci*, 364(1521), 1211–1221. doi:10.1098/rstb.2008.0300 [PubMed: 19528002]
- Graham FK (1978). Constraints on measuring heart rate and period sequentially through real and cardiac time. *Psychophysiology*, 15(5), 492–495. [PubMed: 693763]
- Gruss LF, Langae T, & Keil A (2016). The role of the COMT val158met polymorphism in mediating aversive learning in visual cortex. *Neuroimage*, 125, 633–642. doi:10.1016/j.neuroimage.2015.11.003 [PubMed: 26549298]
- Hamm AO, & Vaitl D (1996). Affective learning: awareness and aversion. *Psychophysiology*, 33(6), 698–710. [PubMed: 8961792]
- Hamm AO, & Weike AI (2005). The neuropsychology of fear learning and fear regulation. *Int J Psychophysiol*, 57(1), 5–14. doi:10.1016/j.ijpsycho.2005.01.006 [PubMed: 15935258]
- Headley DB, & Weinberger NM (2015). Relational associative learning induces cross-modal plasticity in early visual cortex. *Cereb Cortex*, 25(5), 1306–1318. doi:10.1093/cercor/bht325 [PubMed: 24275832]
- Helmholtz H, von. (1866/1962). Concerning the perceptions in general In Trans JS. (Ed.), *Treatise on Physiological Optics* (3rd ed). New York: Dover.
- Hodes RL, Cook EW, 3rd, & Lang PJ (1985). Individual differences in autonomic response: conditioned association or conditioned fear? *Psychophysiology*, 22(5), 545–560. [PubMed: 4048355]
- Insel TR, & Cuthbert BN (2009). Endophenotypes: bridging genomic complexity and disorder heterogeneity. *Biological Psychiatry* 66(11), 988–989. [PubMed: 19900610]
- Junghofer M, Elbert T, Tucker DM, & Rockstroh B (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4), 523–532. [PubMed: 10934911]
- Kamin LJ (1969). Predictability, surprise, attention, and conditioning In Church R & Campbell B (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.
- Kass MD, Rosenthal MC, Pottackal J, & McGann JP (2013). Fear learning enhances neural responses to threat-predictive sensory stimuli. *Science*, 342(6164), 1389–1392. doi:10.1126/science.1244916 [PubMed: 24337299]
- Keil A, Sabatinelli D, Ding M, Lang PJ, Ihssen N, & Heim S (2009). Re-entrant projections modulate visual cortex in affective perception: evidence from Granger causality analysis. *Hum Brain Mapp*, 30(2), 532–540. doi:10.1002/hbm.20521 [PubMed: 18095279]
- Kok P, Jehee JF, & de Lange FP (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270. doi:10.1016/j.neuron.2012.04.034 [PubMed: 22841311]

- Konorski J (1967). Integrative activity of the brain: An interdisciplinary approach. Chicago: University of Chicago Press.
- Lacey JI (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. Paper presented at the Research in Psychotherapy, Washington, DC, USA.
- Lang PJ (1968). Fear reduction and fear behavior: Problems in treating a construct. Paper presented at the Research in Psychotherapy Conference, 3rd, Chicago, IL, USA.
- Lang PJ, Bradley MM, & Cuthbert BN (1998). Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biol Psychiatry*, 44(12), 1248–1263. [PubMed: 9861468]
- Lang PJ, Davis M, & Ohman A (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *J Affect Disord*, 61(3), 137–159. [PubMed: 11163418]
- LeDoux JE (2000). Emotion circuits in the brain. *Annu Rev Neurosci*, 23, 155–184. doi:10.1146/annurev.neuro.23.1.155 [PubMed: 10845062]
- Lipp OV, & Derakshan N (2005). Attentional bias to pictures of fear-relevant animals in a dot probe task. *Emotion*, 5(3), 365–369. [PubMed: 16187873]
- Lissek S, Kaczurkin AN, Rabin S, Geraci M, Pine DS, & Grillon C (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biol Psychiatry*, 75(11), 909–915. doi:10.1016/j.biopsych.2013.07.025 [PubMed: 24001473]
- Lissek S, Powers AS, McClure EB, Phelps EA, Woldehawariat G, Grillon C, & Pine DS (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav Res Ther*, 43(11), 1391–1424. doi:10.1016/j.brat.2004.10.007 [PubMed: 15885654]
- Lovibond PF, & Shanks DR (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *J Exp Psychol Anim Behav Process*, 28(1), 3–26. [PubMed: 11868231]
- McTeague LM, Gruss LF, & Keil A (2015). Aversive learning shapes neuronal orientation tuning in human visual cortex. *Nat Commun*, 6, 7823. doi:10.1038/ncomms8823 [PubMed: 26215466]
- McTeague LM, Laplante M-C, Bulls HW, Shumen JR, Lang PJ, & Keil A (2017). Face perception in social anxiety: visuocortical dynamics reveal propensities for hypervigilance or avoidance. *Biological Psychiatry*, in press.
- McTeague LM, Lang PJ, Laplante M-C, Cuthbert BN, Strauss CC, & Bradley MM (2009). Fearful imagery in social phobia: generalization, comorbidity, and physiological reactivity. *Biological Psychiatry*, 65(5), 374–382. [PubMed: 18996510]
- McTeague LM, Lang PJ, Laplante M-C, Cuthbert BN, Shumen J.R, & Bradley MM (2010). Aversive imagery in posttraumatic stress disorder: trauma recurrence, comorbidity, and physiological reactivity. *Biological Psychiatry*, 67(4), 346–356. [PubMed: 19875104]
- McTeague LM, & Lang PJ (2012). The anxiety spectrum and the reflex physiology of defense: from circumscribed fear to broad distress. *Depress Anxiety*, 29(4), 264–281. doi:10.1002/da.21891 [PubMed: 22511362]
- McTeague LM, Shumen JR, Wieser MJ, Lang PJ, & Keil A (2011). Social vision: sustained perceptual enhancement of affective facial cues in social anxiety. *Neuroimage*, 54(2), 1615–1624. doi: 10.1016/j.neuroimage.2010.08.080 [PubMed: 20832490]
- Miskovic V, & Keil A (2012). Acquired fears reflected in cortical sensory processing: a review of electrophysiological studies of human classical conditioning. *Psychophysiology*, 49(9), 1230–1241. doi:10.1111/j.1469-8986.2012.01398.x [PubMed: 22891639]
- Miskovic V, & Keil A (2013). Visuocortical changes during delay and trace aversive conditioning: evidence from steady-state visual evoked potentials. *Emotion*, 13(3), 554–561. doi:10.1037/a0031323 [PubMed: 23398582]
- Moratti S, & Keil A (2009). Not what you expect: experience but not expectancy predicts conditioned responses in human visual and supplementary cortex. *Cereb Cortex*, 19(12), 2803–2809. doi: 10.1093/cercor/bhp052 [PubMed: 19304914]
- Moratti S, Keil A, & Miller GA (2006). Fear but not awareness predicts enhanced sensory processing in fear conditioning. *Psychophysiology*, 43(2), 216–226. doi:10.1111/j.1464-8986.2006.00386.x [PubMed: 16712592]
- Müller MM, Teder-Salejari W, & Hillyard SA (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nat Neurosci*, 1(7), 631–634. [PubMed: 10196572]

- Myers KM, & Davis M (2002). Behavioral and neural analysis of extinction. *Neuron*, 36(4), 567–584. [PubMed: 12441048]
- Ohman A, & Mineka S (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol Rev*, 108(3), 483–522. [PubMed: 11488376]
- Panitz C, Hermann C, & Mueller EM (2015). Conditioned and extinguished fear modulate functional corticocardiac coupling in humans. *Psychophysiology* 52, 1351–1360. [PubMed: 26189975]
- Pavlov IP (1960). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (6th ed). New York: Dover Publications.
- Perruchet P (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlov J Biol Sci*, 20(4), 163–170. [PubMed: 4069791]
- Petro NM, Gruss LF, Yin S, Huang H, Miskovic V, Ding M, & Keil A (2017). Multimodal Imaging Evidence for a Frontocortical Modulation of Visual Cortex during the Selective Processing of Conditioned Threat. *J Cogn Neurosci*, 1–15. doi:10.1162/jocn_a_01114
- Peck P, De Cesarei A, & Junghofer M (2011). ElectroMagnetoEncephalography software: overview and integration with other EEG/MEG toolboxes. *Comput Intell Neurosci*, 2011, 861705. doi: 10.1155/2011/861705 [PubMed: 21577273]
- Prokasy WF, Grant DA, & Myers NA (1958). Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J Exp Psychol*, 55(3), 242–246. [PubMed: 13513943]
- Regan D (1989). *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*. New York, NY: Elsevier.
- Rescorla RA, & Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement In Black AH & Prokasy WF (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Seligman MEP (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307–320.
- Skrandies W, & Jedynak A (2000). Associative learning in humans--conditioning of sensory-evoked brain activity. *Behav Brain Res*, 107(1–2), 1–8. [PubMed: 10628725]
- Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, & Jacobs GA (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press, Inc.
- Summerfield C, & Egner T (2009). Expectation (and attention) in visual cognition. *Trends Cogn Sci*, 13(9), 403–409. doi:10.1016/j.tics.2009.06.003 [PubMed: 19716752]
- Sutton RS, & Barto AG (1998). *Reinforcement learning: An introduction*. Cambridge, Mass: MIT Press.
- Walker EG (1960). Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J Exp Psychol*, 59, 303–311. [PubMed: 13842503]
- Weymar M, Bradley MM, Hamm AO, & Lang PJ (2013). When fear forms memories: threat of shock and brain potentials during encoding and recognition. *Cortex*, 49(3), 819–826. doi:10.1016/j.cortex.2012.02.012 [PubMed: 22483973]
- Weymar M, Keil A, & Hamm AO (2014). Timing the fearful brain: unspecific hypervigilance and spatial attention in early visual perception. *Soc Cogn Affect Neurosci*, 9(5), 723–729. doi:10.1093/scan/nst044 [PubMed: 23547244]
- Weymar M, Schwabe L, Löw A, & Hamm AO (2012). Stress sensitizes the brain: increased processing of unpleasant pictures after exposure to acute stress. *J Cogn Neurosci*, 24(7), 1511–1518. doi: 10.1162/jocn_a_00174 [PubMed: 22185496]
- Wiedemann G, Pauli P, & Dengler W (2001). A priori expectancy bias in patients with panic disorder. *J Anxiety Disord*, 15(5), 401–412. [PubMed: 11583073]
- Wieser MJ, McTeague LM, & Keil A (2011). Sustained preferential processing of social threat cues: bias without competition? *J Cogn Neurosci*, 23(8), 1973–1986. doi:10.1162/jocn.2010.21566 [PubMed: 20807057]
- Wyart V, Nobre AC, & Summerfield C (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc Natl Acad Sci U S A*, 109(9), 3593–3598. Doi: 10.1073/pnas.1120118109

Highlights:

- We evaluated the impact of threat biases on sensory processing.
- Visuocortical ssVEPs, autonomic and behavioural measures were recorded.
- Initial threat reactivity predicts visuocortical discrimination in extinction.

Figure 1.

Experimental design. Differential classical conditioning paradigm with habituation (H) block and two blocks each of acquisition (A) and extinction (E). Reinforcement schedule is indicated in percentage for each block. Bottom illustrates trial progression for a CS+ trial during the fully reinforced first acquisition block (A1). Repeated Gabor patch within the trial indicates flicker induction for the steady state visually evoked potential. All trials were 7 seconds in duration (dashed gray line) unless the loud noise (UCS) was delivered, which co-terminated with the flickering CS at 8.4 seconds. Acoustic startle probes were presented on two-thirds of the trials at either 4 or 5 seconds into the trial.

Figure 2.

Sympathetic responses to the UCS during acquisition (A1) predict outcome measures in extinction (E1). (a) SCR responses to the UCS broken into tertiary groups for illustrative purposes with UCS onset at 7 sec. (b) Bivariate plots of UCR predicting autonomic, behavioural and visuocortical responses in extinction. (c) Topographies of ssVEP enhancement to CS+ based on tertiary groups from SCR magnitude.

Figure 3. Behavioural expectancy ratings over four experimental blocks, grouped by SCR magnitude in response to the UCS in acquisition (A1). Individuals with the greatest UCS response display greater endorsement of continued UCS expectancy for the CS+ in extinction compared to individuals with smaller skin UCR.

Figure 4.

Heart rate reactivity averaged over extinction, grouped by skin UCR to the UCS in acquisition (A1). Individuals with the greatest skin UCR display enhanced heart rate deceleration to the CS+ in extinction.

Figure 5. Steady state visually evoked potential (ssVEP) signal represented in the time domain (bottom). Gray boxes indicate early and late windows during the trial for the DFT analysis. Disruption of the ssVEP signal from startle probes at 4 and 5 s is clearly visible. Scalp topographies viewed from the back of the head (top) illustrate enhanced responding to the CS+ late during the trial in A1, subsequently shifting in A2 to an enhancement earlier during the trial.

Figure 6. Startle magnitude during all experimental blocks. Average startle responses (a) and single-trial startle responses (b) during all experimental blocks. Conditioning main effects observable across acquisition blocks, as well as A2-E1 blocks.

Figure 7.

Construct validity through the Rescorla-Wagner Learning Rule Model. (a) Two different learning rate values (η) for the CS+ weight vector, reflecting change in associative strength over time. Larger η values reflect more rapid changes over the course of associative learning (contingency tracking), whereas smaller η values are more resilient to extinction after conditioning. (b) Consequently, smaller η values significantly correlated with greater expectancy errors in extinction.

Table 1.

Conditioning effects across neurophysiological measures. Standard error of the mean given in parenthesis.

	A1		A2		E1		E2	
	CS+	CS–	CS+	CS–	CS+	CS–	CS+	CS–
ssVEP (early, μ V)	1.38 (.13)	1.45 (.13)	1.29 (.11)	1.16 (.11)	1.19 (.12)	1.10 (.11)	1.18 (.12)	1.16 (.11)
ssVEP (late, μ V)	.87 (.08)	.72 (.08)	.74 (.07)	.79 (.07)	.75 (.08)	.72 (.07)	.89 (.09)	.79 (.07)
HR dec (bpm)	–6.74 (.43)	–6.14 (.49)	–8.02 (.54)	–7.41 (.47)	–7.79 (.49)	–7.14 (.48)	–7.05 (.47)	–7.58 (.47)
Startle (t scores)	54.93 (.87)	53.04 (.91)	48.15 (.65)	46.50 (.60)	46.10 (.56)	45.48 (.60)	45.76 (.62)	45.53 (.61)
Skin conductance (CS, log μ S)	.02 (.01)	.02 (.01)	.00 (.00)	.00 (.00)	.03 (.01)	.01 (.01)	.01 (.01)	.02 (.01)
Skin conductance (UCS, log μ S)	.17 (.02)	.00 (.00)	.07 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)