Neurocircuitry of Contingency Awareness in Pavlovian Fear Conditioning

Shantanu Madaboosi[1] · Lana Ruvolo Grasser[1] · Asadur Chowdury[1] · Arash Javanbakht[1]

Abstract
In Pavlovian fear conditioning, contingency awareness provides an indicator of explicit fear learning. A less studied aspect of fear-based psychopathologies and their treatment, awareness of learned fear is a common cause of distress in persons with such conditions and is a focus of their treatment. The present work is a substudy of a broader fear-conditioning fMRI study. Following fear conditioning, we identified a subset of individuals who did not exhibit explicit awareness of the CS-US contingency. This prompted an exploratory analysis of differences in "aware" versus "unaware" individuals after fear conditioning. Self-reported expectancies of the CS-US contingency obtained immediately following fear conditioning were used to differentiate the two groups. Results corrected for multiple comparisons indicated significantly greater BOLD signal in the bilateral dlPFC, right vmPFC, bilateral vlPFC, left insula, left hippocampus, and bilateral amygdala for the CS+>CS− contrast in the aware group compared with the unaware group (all p values ≤ 0.004). PPI analysis with a left hippocampal seed indicated stronger coupling with the dlPFC and vmPFC in the aware group compared with the unaware group (all p values ≤ 0.002). Our findings add to our current knowledge of the networks involved in explicit learning and awareness of conditioned fear, with important clinical implications.

Introduction

Associative fear learning involves the formation of memories necessary for adaptive survival and confers the ability to predict and respond to danger through cognitive, autonomic, and psychomotor responses (Liberzon & Ressler, 2012). This results in timely assessment and execution of precautionary behaviors necessary to avoid an impeding threat (Critchley et al., 2002; Goodman, Harnett, & Knight, 2018a; Pellman & Kim, 2015). Although fear learning is traditionally thought of as an automatic process, conscious awareness of learned fear plays an important role in cognitive and behavioral responses to threat, as well as in the psychopathology of fear-based disorders (Goodman, Harnett, Wheelock, et al., 2018b; Liberzon & Abelson, 2016). As such, studying the underlying neurobiological processes involved in awareness of learned fear is important to understand the psychopathology of fear-based disorders.

Pavlovian fear conditioning models fear-based psychopathology in the laboratory setting (Briscione et al., 2014). In the Pavlovian fear conditioning paradigm, a neutral stimulus (conditioned stimulus; CS) is repeatedly paired with an aversive stimulus (unconditioned stimulus; US), until the presentation of the CS alone elicits the conditioned response (CR), which may be defined as an increased skin conductance response (SCR), increased startle response, and/or verbal report of increased fear, anxiety, or expectancy), even in the absence of the US (Carter et al., 2003; Cushman & Fanselow, 2010; Mechias et al., 2010; Weike et al., 2007). While physiological CRs, such as SCR and startle, are traditionally thought of as reflecting implicit fear conditioning, verbal report of increased fear, anxiety, or expectancy are traditionally thought of as reflecting explicit fear conditioning and require higher order cognitive processes (Lovibond, 2019); all characterize responses that can reflect learning and may have some overlap (Lovibond, 2019). The ability to communicate this learned relationship between the CS and US is defined as contingency

✉ Lana Ruvolo Grasser
lgrasser@med.wayne.edu

1 Department of Psychiatry and Behavioral Neurosciences, Wayne State University School of Medicine, 3901 Chrysler Dr., Tolan Park Suite 2C Room 273, Detroit, MI 48201, USA

awareness and serves as an indicator of higher-order cognitive processes in fear learning (Labrenz et al., 2015). Implicit Pavlovian conditioning may occur in the absence of contingency awareness (Balderston & Helmstetter, 2010)—that is the explicit perception that the CS will elicit the US—and explicit Pavlovian conditioning also may occur, representing two potential learning mechanisms (Balderston & Helmstetter, 2010; Carter et al., 2006; Cheng et al., 2003; Knight et al., 2003; Schultz et al., 2013), although this dual-process model is controversial (Lovibond & Shanks, 2002). The extent to which conscious awareness of the CS-US contingency is needed to elicit the associated conditioned response has been a subject of ongoing debate (Labrenz et al., 2015).

Evidence supporting the dual-process model comes from lesion and neuroimaging studies. The amygdala is a predominant structure implicated in fear conditioning and CRs (Kropotov, 2009). Single-patient lesion studies have shown that when the bilateral amygdala are damaged, cognitive awareness of CS-US contingencies is retained; however, CRs are not preserved (Bechara et al., 1995; LaBar et al., 1995). Conversely, patients with bilateral hippocampal damage retain conditioned fear responses but not cognitive awareness of CS-US contingencies (Bechara et al., 1995). Neuroimaging studies echo such findings: two studies have shown greater differential responses in the amygdala and orbito-frontal cortex (OFC) to CS+>CS- in healthy individuals who were deemed aware of the contingencies compared with those who were deemed unaware (Bechara et al., 1995; LaBar et al., 1995). Knight et al. (2009) reported hippocampal and parahippocampal activity in conjunction with contingency awareness; amygdala activity was present regardless of whether contingency awareness was expressed (Knight et al., 2009). The insula, prefrontal cortex (PFC), and dorsal anterior region of the cingulate cortex (dACC) are additional regions widely involved in fear conditioning (Adolphs, 2013; Greco & Liberzon, 2016; Yin et al., 2018). A number of studies attribute the main effect of awareness of CS-US contingencies to multiple regions of interest: the hippocampus, insula, orbitofrontal cortex (OFC), ACC, ventral striatum, parahippocampus, medial frontal gyrus (MFG), and medial PFC (mPFC) (Adolphs, 2013; Baeuchl et al., 2019; Carter et al., 2006; Greco & Liberzon, 2016; Hofmann, 2008; Klucken, Kagerer, et al., 2009a; Klucken, Tabbert, et al., 2009b; Knight et al., 2009; Tabbert et al., 2011; Tabbert et al., 2006; Yin et al., 2018). The functional networks involved in awareness of learned fear are unexplored and knowledge thereof is critical to clinical populations. Fear-based psychopathology comprises both automatic (implicit) and cognitive (explicit) components, yet the majority of research has focused on automatic aspects. More importantly, what brings a patient to the clinic is the subjective awareness of the distress and anxiety linked with the feared cues in

conditions, such as specific phobias, social phobia, and post-traumatic stress disorder (PTSD). This is specifically true in the field of psychiatry where most of the clinical information is gathered through patients' subjective report of distress. For instance, among the important clinical symptoms of patients with PTSD is their subjective distress experienced when facing trauma-related cues. Furthermore, the focus of treatment, and the indicator of improvement, is changes in such subjective reports of distress with encounter with the feared situations and objects of phobias and PTSD.

The present work leverages data from the conditioning phase of an overarching neuroimaging study of fear conditioning and extinction learning. In this study, we used self-reported contingency awareness data to confirm fear learning before continuing to the next phase (extinction learning) of the task. Almost half of our participants (14/31) did not report awareness of the CS-US contingency, prompting us to conduct an exploratory analysis to look at potential differences in brain activity during the conditioning phase in these two groups. The present study investigated the underlying neurobiology that may be associated with differences in cognitively "aware" and "unaware" individuals during fear conditioning (Javanbakht et al., 2017; Javanbakht et al., 2021) via a region of interest (ROI) analytic approach of functional MRI (fMRI) data. We selected an ROI-based approach given established knowledge of areas involved in fear conditioning, as well as our small sample size, to reduce multiple comparisons and chance of Type I error. We hypothesized that the group classified as "aware" of the CS-US contingency would have greater activation in prefrontal regions: dlPFC (Baeuchl et al., 2019), dmPFC (Baeuchl et al., 2019), vlPFC (Kattoor et al., 2013; Lindquist et al., 2012), vmPFC (Phelps, 2004), OFC (Knight & Wood, 2011; Tabbert et al., 2006), and the hippocampus (Baeuchl et al., 2019; Greco & Liberzon, 2016; Klucken, Kagerer, et al., 2009a) compared with the group classified as "unaware" for the CS+ versus the CS−, based on the literature regarding Pavlovian fear conditioning and the explicit learning aspect of fear conditioning. Additionally, given evidence for important role of hippocampus in awareness of learned fear, we hypothesized increased coupling between prefrontal regions, the hippocampus, and the regions associated with threat detection (amygdala, insula, and dACC) in the aware group compared with the unaware group for the CS+ versus the CS− during conditioning (Phelps et al., 2004; Sridharan et al., 2008; Thomson & Jaque, 2017; Uddin, 2016).

Materials and Methods

The present study encompasses an exploratory analysis of fear conditioning data from an fMRI fear conditioning study, prompted by findings of a subgroup of participants who did

not report awareness of the CS-US contingency. Data from the extinction and extinction recall phases will not be presented in this publication; however, the paradigm and results from those phases are described in Javanbakht et al. (2021).

Participants

Thirty-seven (37) healthy male and female (n_{female} = 20) participants aged 18 to 45 ($\bar{x}_{age}$ = 26.18, SD_{age} = 4.61) were recruited for an fMRI study of fear conditioning and extinction learning using approved flyers and university forum posts. All study procedures described herein were approved by the Institutional Review Board at Wayne State University (IRB#012316B3F). Oral consent and initial eligibility screening were completed via phone interview. Exclusion criteria included: current psychiatric illness (except for specific phobias and history of substance related disorders more than 1 year prior), neurological impairment that may affect normal brain function, pregnancy, metal in body, history of significant closed head injury, or claustrophobia. Those who met initial criteria then came to the lab to provide written informed consent and complete an in-person interview. Participants also were presented with the 95dB white noise burst (the US) through a pair of noise-cancelling headphones to ensure that they were able to hear and tolerate the sound. Brain scanning was completed on a subsequent day. Six participants were excluded due to excess motion (>3 mm) during MR imaging, according to standard guidelines consistently applied in our lab and others (Evans et al., 2016; Javanbakht et al., 2021; Javanbakht et al., 2016; Javanbakht et al., 2015). This left a total of n = 31 (n_{male} = 16, n_{female} = 15, $\bar{x}_{age}$ = 26.18, SD_{age} = 4.78) participants with usable data for analyses (Table 1).

Summary statistics of the distributions shown in the graph. Racial/ethnic data were not collected. No participants indicated presence of a specific phobia or history of substance use disorder greater than 1 year prior.

Assessments

During the in-person interview, the Mini International Neuropsychiatric Interview (MINI) was used to screen participants for possible psychiatric disorders and confirm self-reported information provided during the phone interview (Sheehan et al., 1998). Based on the MINI, included participants did not report any clinically significant anxiety symptoms, nor any other clinically significant psychiatric symptoms. None of the participants screened positive for specific phobias or history of substance related disorders.

Image Acquisition and Paradigm

Structural and functional MRI data were collected for this study on a 3T Siemens Verio system with a 32-channel volume head coil. BOLD signal was obtained during the fear conditioning phase of the paradigm (as well as the fear extinction phase which is not reported on in this manuscript). A high-resolution (1 mm³) structural T_1-weighted anatomical image was first collected (3D Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence, TR = 2,150 ms, TE = 3.53 ms, TI = 1,100 ms, flip-angle = 8°, FOV = 256 x 256 x 160 mm³, 160 axial slices of thickness = 1 mm, pixel resolution = 1 x 1 x 1 mm³, and TA = 4:59 minutes). Functional BOLD signal acquisition followed (multiband gradient Echo EPI fMRI, 310 vol, TR = 2 s, TE = 29 ms, multiband factor = 3, FOV = 256 x 256 x 144 mm³, acquisition matrix = 128 x 128, 72 axial slices, pixel resolution = 2 x 2 x 2 mm³, and TA = 10:48). MRI technicians continuously tracked signal acquisition for quality control in real time, and no issues of poor signal were noted. Continuous monitoring of participant behavior did not indicate any participants sleeping. After verbal inquiry, no participants responded that they fell asleep during scanning.

A Dell fMRI control and integration system ran the paradigm using E-Prime 2.0. A 3T MRI compatible projector displayed the paradigm on the projector screen and a hi-fidelity audio system played any sound that may have been presented as part of the task through MR compatible headphones worn by participants. The Pavlovian fear conditioning paradigm used in this study has been extensively used to study fear learning (Marin et al., 2017; Milad et al., 2008; Milad et al., 2005; Milad et al., 2009; Rabinak et al., 2017; Raij et al., 2018). Images of three different colored lamps were used as conditioned stimuli (CS1, CS2, and CS3) (Fig. 1).

Before conditioning, participants were exposed to all three images that would be shown in latter phases. During this habituation phase, participants were first presented with four trials of each CS and told they would not hear any noises. Fear conditioning immediately followed, at which time participants were told that they would see the previous images and informed that some of those images may be paired with a loud noise. The US (the loud noise) was presented 3.5 seconds following onset of the CS+s (either CS1 or CS2; the third CS, CS3, was never paired with the US, rendering it the CS−). The US was a 95 dB x 500 ms white noise burst played through the

Table 1 Distribution of age for participants included in this study separated by group and gender

Gender	Group	Age				
		n	$\bar{x}$	sd	Min	Max
Male	Aware	10	27.46	6.64	21.29	45.43
	Unaware	6	26.82	5.03	22.49	35.37
Female	Aware	7	24.83	1.26	22.96	26.55
	Unaware	8	25.25	4.05	21.46	31.80

MR-compatible noise cancelling headphones (Sperl et al., 2016) which co-terminated with the CS+s. The conditioning phase contained 15 trials, during which each of the three lamp colors would be cycled through in random order without replacement (45 images shown in total, shown for 4 seconds each with jittered inter-trial intervals between 6 and 12 seconds). The US was paired with the CS+s at a 66% reinforcement rate. Lamp colors assigned to CS1, CS2, and CS3 were randomized and counterbalanced across participants (Javanbakht et al., 2017; Javanbakht et al., 2021). No participants were given additional nor differential information regarding the CS-US contingencies before or during the condition phase.

Contingency Awareness

Immediately upon completion of conditioning, participants were presented with an image of each CS and concurrently asked to verbally rate how much they expected to hear the loud noise when presented with each CS as shown in Fig. 2.

Ratings were made on a 5-point Likert scale: 1 = "Not at All" and 5 = "Very Much So" (Javanbakht et al., 2017; Javanbakht et al., 2021; Tabbert et al., 2006). This verbally reported expectancy data was acquired for both CS+s (CS1 and CS2), as well as the CS− (CS3). The expectancy ratings for the CS+s were averaged, and if the average expectancy rating for the CS+s was greater than that of the expectancy rating for the CS-, then participants were classified as "aware" (Javanbakht et al., 2017; Javanbakht et al., 2021; Tabbert et al., 2006). If the average expectancy rating for the CS+s was equal to or less than that of the expectancy rating for the CS−, then participants were classified as "unaware" (Javanbakht et al., 2017; Javanbakht et al., 2021; Tabbert et al., 2006). See Eq. 1 for a representation of the calculation. A repeated measures ANOVA was conducted to compare

mean expectancy responses (the dependent variable) by CS (CS+ vs. CS−; within-subjects effect) and group (aware vs. unaware; between-subjects effect). Type III sum of squares was applied to account for the unbalanced group sizes. Tukey's HSD post-hoc was then used to determine significant interaction effects of CS by group on expectancies.

$$\begin{array}{ll} Group\ Expectancies \\ Awareness = \overline{CS+} > CS- \\ Uawareness = \overline{CS+} \leq CS- \end{array} \qquad (1)$$

Equation used to determine if participants were aware or unaware of the CS-US contingency. If the average of the two ratings for CS+1 and CS+2 was greater than that of the rating for the CS−, then the individual was classified as "aware" of the CS-US contingency. If the average of the two ratings for CS+1 and CS+2 was less than or equal to that of the rating for the CS−, then the individual was classified as "unaware" of the CS-US contingency. So, a person who was categorized as aware would be one who accurately predicts the loud noise more with the CS+1 and CS+2 than the CS−. Those whose difference on contingency was equal to 0 were categorized as unaware.

Image Processing

fMRI data were pre-processed and analyzed using MATLAB R2019a (MathWorks, 2013; 2019) with the Statistical Parametric Mapping toolbox (SPM12) (Wellcome Centre for Human Neuroimaging, 2020). Imaging data from all 15 trials were included for each CS.

Pre-processing For each participant, manual AC-PC alignment was performed on the T$_1$ structural scan, and the corresponding reorientation matrix was then applied to the 310 EPI

*Colors assigned to each conditioned stimulus (CS) is randomized between participants.
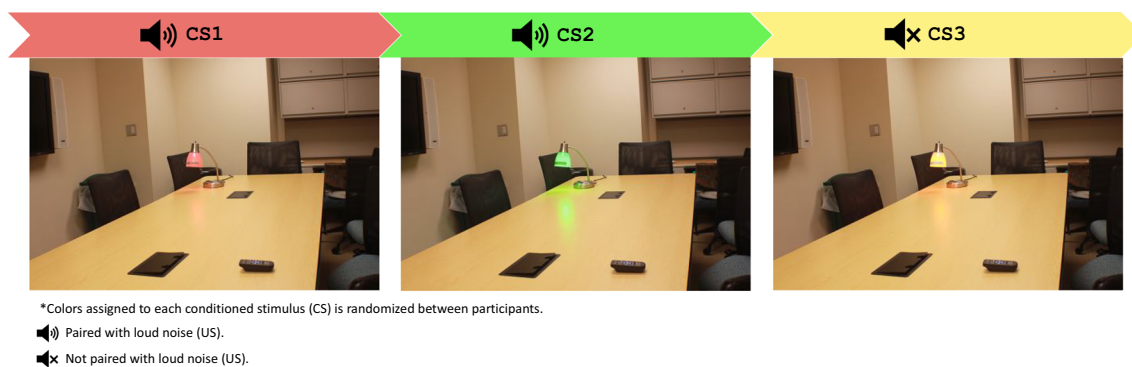🔊 Paired with loud noise (US).
🔇 Not paired with loud noise (US).

Fig. 1 Paradigm shown to participants during conditioning. Two of the three lamp colors were paired with the loud noise (US), represented by the "Volume on" symbol, to form the CS+. The remaining lamp color was unpaired, the CS− (CS3), denoted by the "Volume off" symbol. Lamp colors were randomized to CS+1, CS+2, and CS− across participants, but remain constant within each subject throughout the experiment. Each image is shown for 4 seconds per trial with a total of 15 trials; all 15

trials were including for imaging analyses. Noise onsets for the 10 trials of CS+1 / CS+2 that are reinforced occur 3.5 seconds after image onset with a duration of 0.5 seconds, such that stimuli co-terminate. The time between image offsets and the next image onset (inter-trial interval; ITI) is temporally separated by a fixation cross. ITI durations are randomly selected, without replacement, from options of 6, 8, 10, or 12 seconds.
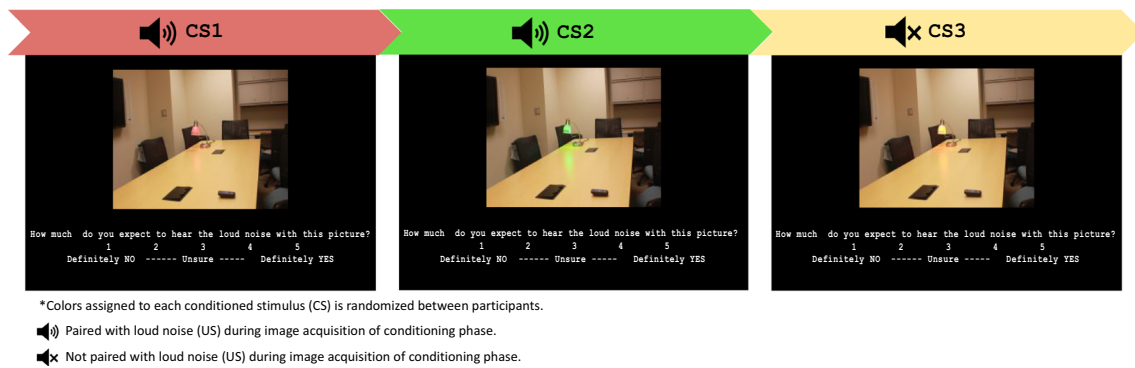
*Colors assigned to each conditioned stimulus (CS) is randomized between participants.

🔊 Paired with loud noise (US) during image acquisition of conditioning phase.

🔇 Not paired with loud noise (US) during image acquisition of conditioning phase.

Fig. 2 How contingency responses were collected. Prompts that participants saw when asked for expectancy ratings of each CS. Ratings were on a scale from one through five; one indicated the loud noise was not expected at all, and five indicated the noise was heavily anticipated for the image. This served as an indicator of cognitive awareness of CS-US contingencies; those who are aware of the CS would show high expectancy for the images that have been paired with the loud noise during the conditioning phase.

images. Functional images were then corrected to the actual differences in acquisition time between slices, realigned to the first image, and co-registered to the subject's T_1 structural scan. The T_1 image was segmented and bias correction for scanner noise was performed as a standard component of our processing pipeline (Javanbakht et al., 2021). The T_1 image was then spatially normalized to MNI template with the resultant deformations applied to the realigned and co-registered images. Then, the EPI images were spatially smoothed using a Gaussian filter (8 mm, full width half maximum).

First Level Onset times, durations, and event names were logged at the time of acquisition and extracted from E-Prime outputted log files. This information was used to construct a general linear model (GLM) for blood oxygen level dependency (BOLD) that is predictive of the image being shown at each time point. To account for serial correlation, an autoregressive model was used; a canonical hemodynamic reference waveform was used to convolve box car modelled regressor vectors corresponding to each event type. Included in the effect of no interest were the six motion parameters (translational = [xx, yy, zz], rotational = [roll, pitch, yaw]). T-weighted contrasts for CS+ BOLD signal intensities greater than CS− BOLD signals intensities (CS+>CS−) were created for each participant. This represented the overall fluctuations in BOLD signal intensities attributed to CS+ minus the overall fluctuations in BOLD signal intensities attributed to CS−. This model was later used for second-level group analysis.

Second Level ROI-based analysis informed by the literature regarding Pavlovian fear conditioning, and more specifically the explicit learning aspect of fear conditioning, was performed (Adolphs, 2013; Baeuchl et al., 2019; Bechara et al., 1995; Carter et al., 2006; Greco & Liberzon, 2016; Klucken, Kagerer, et al., 2009; Klucken, Tabbert, et al., 2009; Tabbert et al., 2011; Tabbert et al., 2006). Bilateral masks for each *a*

priori defined region of interest (ROI) were created separately using WFU_PickAtlas (Maldjian et al., 2003). ROIs were masked separately due to the disproportionality in sizes between each region. Combining ROIs into one mask would result in cluster size thresholds that are physiologically impossible for smaller regions to meet after correcting for multiple comparisons (i.e., minimum cluster thresholds larger than entire regions such as the amygdala and hippocampus). The Automated Anatomical Labelling (AAL) software package and digital atlas of the human brain was used to define masks for the amygdala, insula, dACC, and hippocampus (Tzourio-Mazoyer et al., 2002). Brodmann Areas (BA) were used to define masks for prefrontal areas (see Supplementary Figure 1). BA 8, 9, and 46 are associated with the dlPFC (Mylius et al., 2013); voxels within these areas, excluding any voxels within x = 0 ± 25 mm, were masked to isolate lateral parcellations from overlaps with the dmPFC. BA 8, 9, and 32 are associated with the dmPFC (Kober et al., 2008; Watanabe, 2017); voxels that did not intersect with a 30-mm radius centered at [0, 44, 26] were excluded in this mask to distinguish medial parcellations from overlaps with the dlPFC. BA 12, 44, 45, and 47 are associated with the vlPFC (Lévesque et al., 2003; R. C. O'Reilly, 2010); all voxels within these areas were include in the mask. BA 10, 14, 25, and 32 are associated with the vmPFC and OFC (McNamara et al., 2008; Qiu et al., 2014; Saez et al., 2018); voxels that did not intersect with a 45-mm radius centered at [0, 56, −28] were excluded in this mask to distinguish medial parcellations from overlaps with the lateral parcellations (LaBar et al., 1995).

A two-sample t-test was implemented to compare differences in BOLD signal intensities between the aware and unaware participants, for the CS+>CS− contrast created at first level. BOLD signal intensities were also explored for each group separately and analyzed using the same procedures. Each ROI's intensity map was then Markov chain Monte Carlo (MCMC) simulated using the 3dClustSim function of AFNI's AlphaSim software. 3dClustSim calculates the

minimum cluster size needed to signify true detection of activation as opposed to noise at $p < 0.01$ (Friedman et al., 2017; Ward, 2000). Thresholds are computationally calculated probabilities of random field noise; true activation clusters are denoted by contiguous voxel intensities that exceed these thresholds. Clusters that did not meet these thresholds were excluded as noise. Calculated thresholds are provided in Supplementary Table 1.

Psychophysiological Interactions Psychophysiological interaction (PPI) analysis is a method of analysis that gives insight into how a single voxel and its surrounding voxels, located at a specified ROI, fluctuate, regressed to the main effect of a task (in our case CS+>CS−) (Casey et al., 1995; Friston et al., 1997; Ladouceur et al., 2011). This allows for observations of how one region may be intrinsically connected with another when confronted with the same task. We chose to use standard PPI to understand differential network functionality during a condition—here, contingency awareness (Friedman et al., 2017; Friston et al., 1997; O'Reilly et al., 2012; Silverstein et al., 2016). Because of the consistent findings from previous studies and ours about the role of hippocampus in awareness of learned fear (Bechara et al., 1995; Phelps, 2004; Sanders et al., 2003; Squire, 1992), we used a seed voxel from the left hippocampus to examine its connectivity with other areas involved in learning and awareness of conditioned fear. Timeseries values were extracted and averaged across all voxels within a 4-mm sphere centering at the voxel located on the left hippocampus [−24 −33 −3]. This coordinate was selected as the peak activation region from the ROI-based analyses (Supplementary Figure 3). A whole brain PPI analysis was conducted and resulting activation maps were masked using the same masks mentioned in our original analysis. Each ROI was MCMC simulated and analyzed using the exact same methods as the aforementioned regional activation analysis. The βs were calculated for each voxel and corrected for the main effect of CS+>CS− BOLD signal intensities. Significant clusters are those that were selected using the previously defined masks (amygdala, insula, dACC, hippocampus, dmPFC, dlPFC, vlPFC, vmPFC, and OFC) and show some level of covariance in intensities during task. The differences in the covariance between the left hippocampus and all other masked regions is then compared for the aware group, unaware group and aware versus group.

For both regional activation and PPI analyses, we used a Bonferroni correction for 8 ROIs with an original alpha level of 0.01. This yielded a corrected alpha level of 0.00125.

Results

A total of 37 participants completed the fear conditioning task, and 6 were excluded due to excessive motion (>3 mm),

leaving a total of 31 participants with usable data. Based on Eq. 1, the aware group consisted of 17 participants ($n_{female} = 7$, $\bar{x}_{age} = 26.38$, $SD_{age} = 5.21$); the unaware group consisted of 14 participants ($n_{female} = 8$, $\bar{x}_{age} = 25.93$, $SD_{age} = 4.38$).

Expectancy Rating Differences

A repeated measures ANOVA was used to compare expectancy ratings by group and CS. This showed a large significant main effect of CS ($F(1,29) = 117$, $p < 0.001$, $\eta_p^2 = 0.801$), a small significant main effect of group ($F(1,29) = 6.11$, $p = 0.020$, $\eta_p^2 = 0.174$), and a significant large interaction effect of CS and group ($F(1,29) = 198$, $p < 0.001$, $\eta_p^2 = 0.872$; Table 2). Post-hoc tests revealed mean expectancy ratings to be significantly different for four distinct comparisons (Fig. 3). CS+ ratings were significantly greater than CS− ratings in the aware group at $p < 0.001$. Significantly greater expectancy rating were observed for CS+ in the aware group compared with CS+ in the unaware group at $p = 0.005$. Significantly greater mean expectancy rating were observed for CS+ in the unaware group compared with CS− in the aware group at $p < 0.001$. Lower mean expectancy ratings were observed for CS− in the aware group compared to CS− in the unaware group at $p < 0.001$, suggesting potential lack of discrimination in the unaware group.

Regional Differences in BOLD Signal Intensities

Region of interest (ROI) based results that are summarized in Fig. 4 were derived using a two-sample t-test of BOLD signal intensity differences between the aware group and unaware group for BOLD signal intensities greater in the CS+ condition than the CS− condition. Among the *a priori* specified ROIs, BOLD signal intensity was significantly greater in the bilateral dlPFC, right vmPFC, bilateral vlPFC, left insula, left hippocampus, and bilateral amygdala for the aware group compared with the unaware group when confronted with the same task (CS+>CS−), Markov chain Monte Carlo (MCMC) minimum cluster threshold corrected $p < 0.01$. Response parameters are plotted by group in Supplementary Figure 2 and indicate responses to be generally higher for the aware group compared with the unaware group. In the aware group alone, BOLD signal intensity was significant in the bilateral dlPFC, bilateral dmPFC, bilateral vmPFC, bilateral vlPFC, bilateral insula, and right amygdala, MCMC minimum cluster threshold corrected at $p < 0.01$ (Fig. 5). No significant differences were found in the reverse comparison (unaware > aware) or for the unaware group alone.

Psychophysiological Interactions

A seed that encompassed all voxels within a 4mm radius centered at [−24, −33, −3]—left hippocampus—was convolved

Table 2 Repeated measures ANOVA results comparing expectancy ratings by CS and group

Within-subjects effects

	Sum of squares	df	Mean square	F		p	η^2_p
CS	21.71	1	21.710	117		<.001	0.801
CS * Group	36.71	1	36.710	198		<0.001	0.872
Residual	5.39	29	0.186				

Between-subjects effects

	Sum of squares	df	Mean square	F		p	η^2_p
Group	5.54	1	5.543	6.11		0.020	0.174
Residual	26.33	29	0.908				

Post-hoc comparisons - Condition ✳ Group

Comparison					Mean difference	SE	df	t	p_{tukey}
CS	Group		CS	Group					
CS+	Aware	-	CS+	Unaware	0.945	0.267	40.4	3.54	0.005
		-	CS−	Aware	2.735	0.148	29.0	18.50	<0.001
		-	CS−	Unaware	0.588	0.267	40.4	2.20	0.139
	Unaware	-	CS−	Aware	1.790	0.267	40.4	6.71	<0.001
		-	CS−	Unaware	-0.357	0.163	29.0	-2.19	0.149
CS-	Aware	-	CS−	Unaware	-2.147	0.267	40.4	-8.05	<0.001

Note. Type 3 Sums of Squares

Tukey's HSD post-hoc test shows where significant differences exist. This analysis revealed a significant main effect of group and CS, as well as a significant interaction effect of group*CS.

to identify spontaneous fluctuations over time (Supplementary Figure 3).

A subsequently generated time series was used to construct a GLM, with βs corrected for the main effect of task (areas of intensity during CS+ trials greater than CS− trials). BOLD signal intensities in *a priori* specified ROIs (the same as those defined for the original analysis above) coupled with left hippocampus were then compared. A two-sample *t*-test

showed the left hippocampus more strongly coupled with the right dlPFC and right vmPFC during CS+>CS− in aware versus unaware participants (Fig. 6; note that the vmPFC did not survive Bonferroni correction for multiple ROIs). Response parameters are plotted by group in Supplementary Figure 2. Figs. 7 and 8 indicate left hippocampal coupling in each group (aware and unaware) separately.
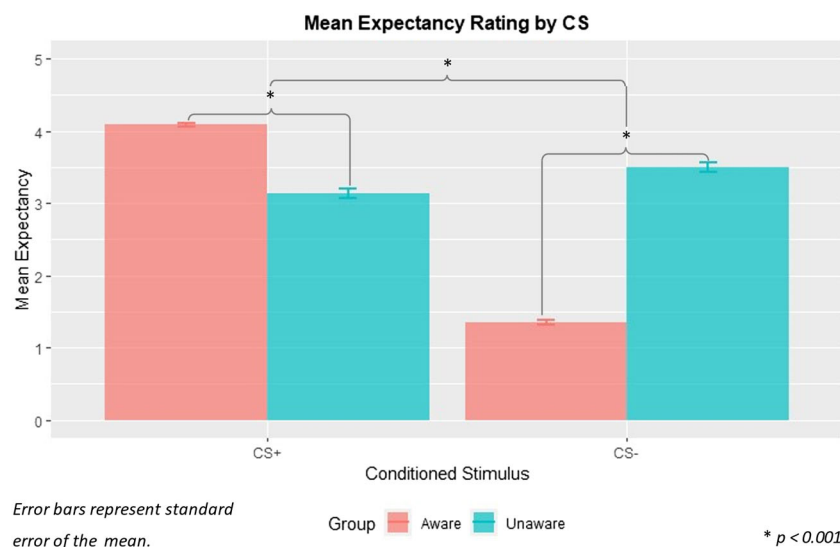

Mean Expectancy Rating by CS

Error bars represent standard error of the mean.

Group: Aware, Unaware

* p < 0.001

Fig. 3 Expectancy differences by condition and between groups. Mean expectancy ratings for CS+US and CS− by group. *Significant differences identified using a RMANOVA with Tukey HSD post-hoc. Significant differences are shown between groups for CS ratings as well as within condition by group. CS+ ratings did not significantly differ from the CS− rating for the unaware group, t(29) = −2.19, p = 0.149, but did significantly differ for the aware group, t(29) = 18.50, p < 0.001.
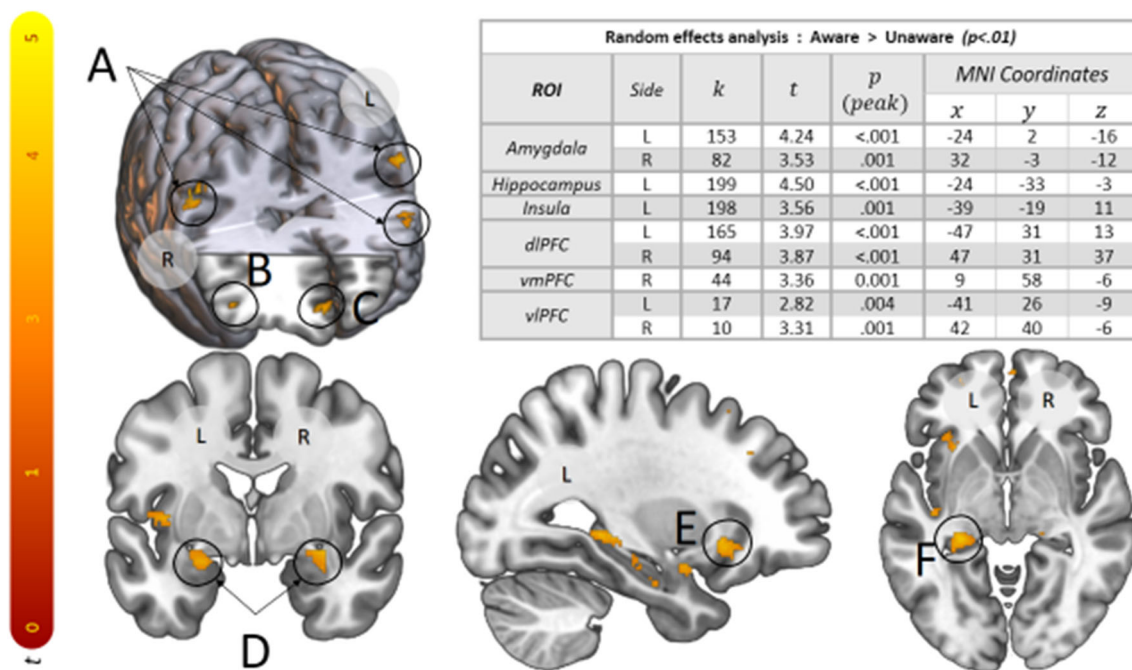
Random effects analysis : Aware > Unaware *(p<.01)*

ROI	Side	k	t	p (peak)	MNI Coordinates x	y	z
Amygdala	L	153	4.24	<.001	-24	2	-16
	R	82	3.53	.001	32	-3	-12
Hippocampus	L	199	4.50	<.001	-24	-33	-3
Insula	L	198	3.56	.001	-39	-19	11
dlPFC	L	165	3.97	<.001	-47	31	13
	R	94	3.87	<.001	47	31	37
vmPFC	R	44	3.36	0.001	9	58	-6
vlPFC	L	17	2.82	.004	-41	26	-9
	R	10	3.31	.001	42	40	-6

Fig. 4 Two-sample t-test of intensity differences comparing aware vs. unaware for CS+>CS−. BOLD signal intensities significantly greater in the aware group compared with the unaware group for masked regions, using a two-sample t-test, are shown in the table above. Each region displayed survived the Monte-Carlo simulated cluster threshold needed to signify nonrandom activation at p threshold <0.01. Areas that showed greater intensities for the aware group were the (**A**) bilateral dlPFC, (**B**) bilateral vlPFC, (**C**) right vmPFC, (**D**) bilateral amygdala, (**E**) left Insula, and (**F**) left hippocampus. All regions displayed survived a Bonferonni correction for multiple ROIs (t ≥ 3.2272, p ≤ 0.00125) except for the left vlPFC

Random effects analysis : Aware *(p<.01)*

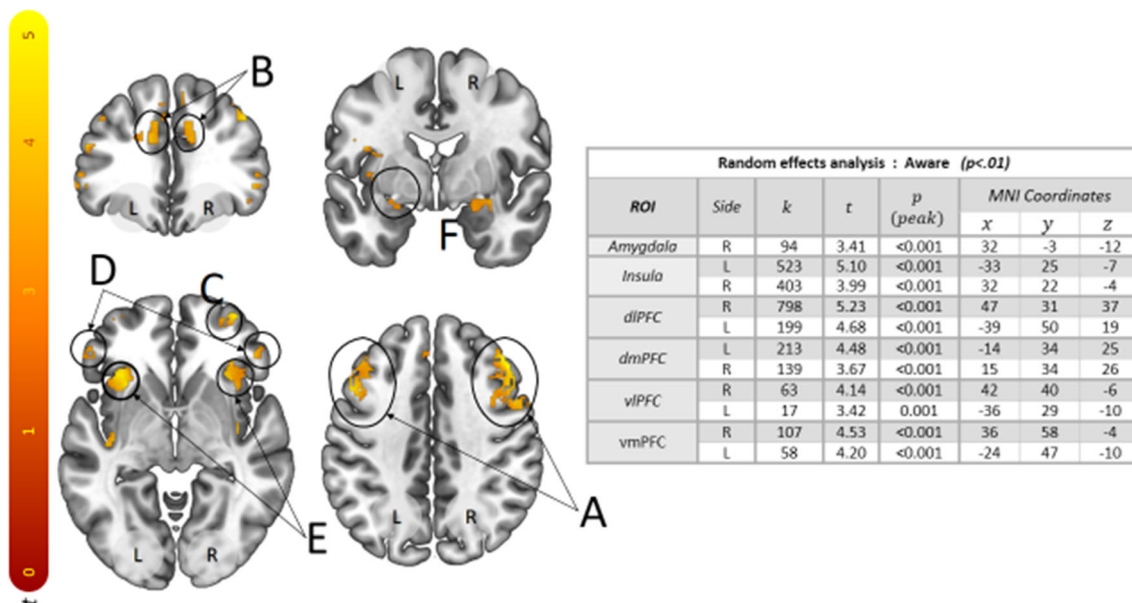| ROI	Side	k	t	p (peak)	MNI Coordinates x	y	z
Amygdala	R	94	3.41	<0.001	32	-3	-12
Insula	L	523	5.10	<0.001	-33	25	-7
	R	403	3.99	<0.001	32	22	-4
dlPFC	R	798	5.23	<0.001	47	31	37
	L	199	4.68	<0.001	-39	50	19
dmPFC	L	213	4.48	<0.001	-14	34	25
	R	139	3.67	<0.001	15	34	26
vlPFC	R	63	4.14	<0.001	42	40	-6
	L	17	3.42	0.001	-36	29	-10
vmPFC	R	107	4.53	<0.001	36	58	-4
	L	58	4.20	<0.001	-24	47	-10

Fig. 5 Random effects analysis of BOLD signal intensities in aware group for CS+>CS−. BOLD signal intensities significantly greater in the aware group alone for CS+>CS− are shown in the table above. Each region displayed survived the Monte-Carlo simulated cluster threshold needed to signify non-random activation at p threshold <0.01. Areas that showed significant intensities for the aware group were the (**A**) bilateral dlPFC, (**B**) bilateral dmPFC (left not shown), (**C**) bilateral vmPFC, (**D**) bilateral vlPFC, (**E**) bilateral Insula, and (**F**) right amygdala. All regions displayed survived a Bonferonni correction for multiple ROIs (t ≥ 3.2272, p ≤ 0.00125)

Discussion

In this study, we investigated the neurocircuitry of contingency awareness in Pavlovian fear conditioning. Behavioral and electrophysiological studies have shown evidence for successful fear conditioning with and without cognitive awareness of CS-US contingencies (Tabbert et al., 2011; Weike et al., 2007). Given that almost half of participants in a fear conditioning and extinction learning study did not present with cognitive awareness of the CS-US contingency (Javanbakht et al., 2021), we leveraged these data to build on the growing literature regarding explicit learning in Pavlovian fear conditioning. We hypothesized that those who developed awareness of the CS-US contingency would show greater responsivity in the prefrontal region (including the dlPFC, dmPFC, vlPFC, and vmPFC/OFC) due to its indications for higher order cognitive processing, including attention, working memory, task execution, and emotion regulation. We also hypothesized greater hippocampal activity in the aware group due to the hippocampus's involvement in emotional memory formation and contextual emotion regulation (Dere et al., 2010).

Significantly greater BOLD signal intensity was observed in the amygdala and insula during fear conditioning for CS+>CS− in the aware group compared to the unaware group. We did not expect to see any significant differences in the amygdala based on previous work on awareness wherein this structure is predominantly implicated in implicit (autonomic/unconscious) learning (Carter et al., 2006). Yet, we masked for this region due to its omnipresence in fear conditioning, anxiety, and attention (Davis, 1997). Nevertheless, amygdala intensities were observed along with the insula, both of which are components of the salience network (SN): a network that generates autonomic responses to threat-related cues in the environment (Selemon et al., 2019). The anterior insula is implicated in emotional awareness—integrating motivational information and sensory information to generate outcomes (Gu et al., 2013; Picard, 2013). The SN has widespread connections throughout the brain, including executive networks, which may influence behavioral responses to salient cues. Increased salience detection and formation of conditioned fear memories, which is attributed to the amygdala, may facilitate cognitive awareness of the CS-US contingencies.

Consistent with our initial hypothesis, the aware group showed greater signal intensity in the dlPFC, vmPFC, and vlPFC compared with the unaware group, which provides support towards the role of the central executive network (CEN) in cognitive awareness of learned fear (Adamson et al., 1999; Cocchi et al., 2012). The CEN is a large-scale intrinsic brain network that supports higher-order cognitive processes, including attention, working memory, and task execution. The dorsolateral prefrontal cortex (dlPFC) is associated with working memory and selective attention (Curtis & D'Esposito, 2003). In tandem with the ventrolateral prefrontal
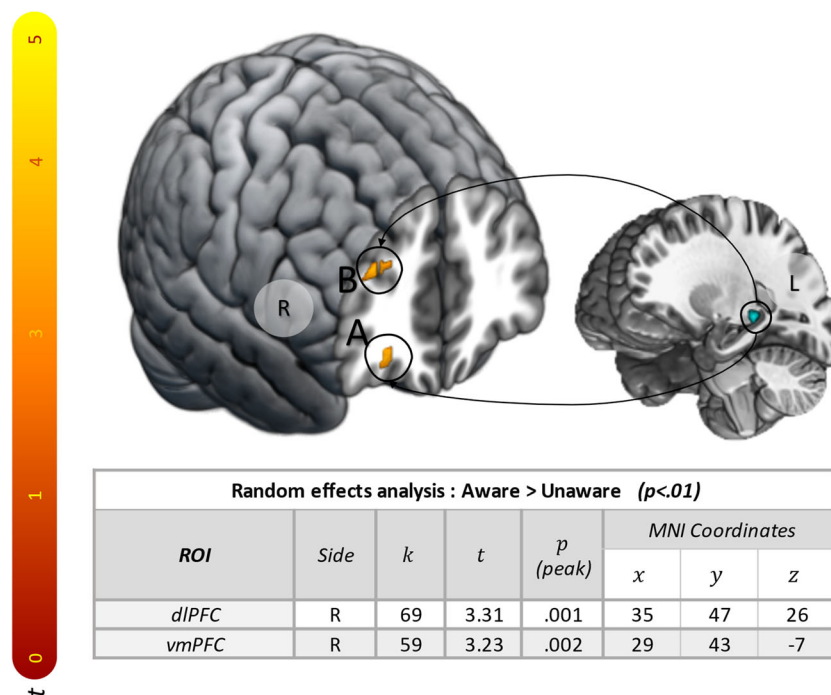

Fig. 6 Psychophysiological interactions between left hippocampus and regions of interest's intensity in aware versus unaware for CS+>CS−. *Psychophysiological interactions indicated that BOLD signal in the dlPFC (**A**) and vmPFC (**B**) was functionally coupled with the left hippocampus (seed region) for the aware group compared to the unaware group, CS+>CS−. The dlPFC survived a Bonferonni correction for multiple ROIs (t ≥ 3.2272, p ≤ 0.00125); however, the vmPFC did not*

	Random effects analysis : Aware > Unaware *(p<.01)*						
					MNI Coordinates
ROI	*Side*	*k*	*t*	*p (peak)*	*x*	*y*	*z*
dlPFC	R	69	3.31	.001	35	47	26
vmPFC	R	59	3.23	.002	29	43	-7

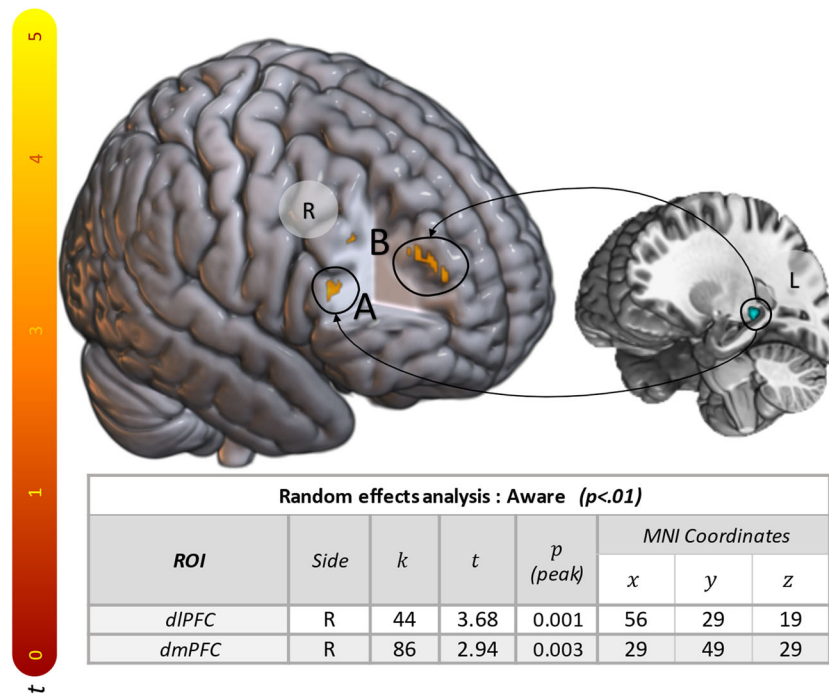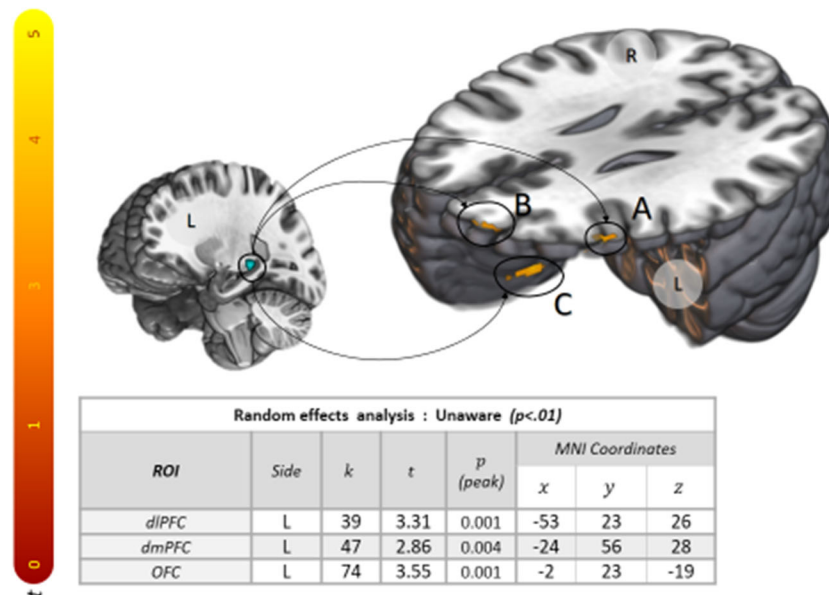| Random effects analysis : Aware (p<.01)							
ROI	Side	k	t	p (peak)	MNI Coordinates		
					x	y	z
dlPFC	R	44	3.68	0.001	56	29	19
dmPFC	R	86	2.94	0.003	29	49	29

Fig. 7 Psychophysiological interactions between left hippocampus and regions of interest's intensity in aware group for CS+>CS−. Psychophysiological interactions between the left hippocampus and ROIs, showed significant co-activation for the aware group in the (**A**) dlPFC and (**B**) dmPFC. The dlPFC survived a Bonferonni correction for multiple ROIs t ≥ 3.2272, (p ≤ 0.00125); however, the dmPFC did not

cortex (vlPFC), the dlPFC is involved in emotion regulation and has an additional role in threat appraisal (Rilling & Sanfey, 2009; Silvers et al., 2015; Staudinger et al., 2011; Sturm et al., 2016). Such increased activation in executive regions observed in the aware group may have multiple implications for awareness, including conscious decision making, reasoning, working memory, inhibition, outcome prediction, and selective attention. Consistent with previous studies, no significant activation was found for the reverse comparison (unaware > aware) during the task (CS+>CS−) (Klucken, Kagerer, et al., 2009; Klucken, Tabbert, et al., 2009; Tabbert et al., 2011).

Random effects analysis : Unaware (p<.01)							
ROI	Side	k	t	p (peak)	MNI Coordinates		
					x	y	z
dlPFC	L	39	3.31	0.001	-53	23	26
dmPFC	L	47	2.86	0.004	-24	56	28
OFC	L	74	3.55	0.001	-2	23	-19

Fig. 8 Psychophysiological interactions between left hippocampus and regions of interest's intensity in unaware group for CS+>CS−. Psychophysiological interactions between the left hippocampus and ROIs showed significant co-activation for the unaware group in the (**A**) left dlPFC, (**B**) left dmPFC, and (**C**) left OFC. The dlPFC and OFC survived a Bonferonni correction for multiple ROIs (t ≥ 3.2272, p ≤ 0.00125); however, the dmPFC did not

Additionally, we observed greater hippocampal signal intensity in the aware group compared with the unaware group. We have previously suggested that contingency awareness can function as a component of context, leading to activation of the hippocampus (Javanbakht et al., 2017). Indeed, lesions of the hippocampus are associated with a loss of cognitive awareness of learned fear, whereas conditioned fear responses are preserved, indicating that the hippocampus may be necessary for cognitive awareness (Bechara et al., 1995). Due to the significant role the hippocampus plays in this task and because of our ROI-based findings, the hippocampus was set as a seed for an exploratory PPI analysis. To our knowledge, is the first study to examine the functional connectivity using PPI associated with contingency awareness in Pavlovian fear conditioning. Involved in higher-order cognition and facilitation of context-dependent fear and safety learning, the hippocampus plays a role in episodic, emotional, and spatial memory formation–conducive to the ability to verbally express learned awareness of the CS-US contingency (Hamm & Weike, 2005; Leuchs et al., 2019; Sevenster et al., 2014). We showed that the left hippocampus was more strongly coupled with the right dlPFC and right vmPFC for the aware group versus the unaware group (although the vmPFC cluster did not survive Bonferroni correction for multiple ROIs). Traditionally, the hippocampus is characterized as a component of the default mode network (DMN) and may provide a link between the CEN and the SN to integrate salient cues, cognitive modulation of cues, and result in memory formation required for learned awareness.

These findings have important clinical implications. Especially in the field of psychiatry, what brings a patient to the clinic is the subjective experience of distress—namely, awareness of fear when exposed to feared objects and situations for persons with specific phobias; awareness of fear when exposed to reminders of a traumatic event for persons with PTSD. Laboratory models of phobia and PTSD are based in fear conditioning, and their focus has mostly been on the automatic/implicit aspects of fear learning and responses. Studies, such as the present work, can contribute to expanding use of these models to provide insight into the higher-level cognitive aspects of fear and its learning, which often is what distresses the patient. Also, in clinical work, improvement by treatment often is assessed by changes in such subjective levels of distress reported by patients and is the target of such treatment. Expanding the study of awareness of fear learning and its neurobiology can provide pivotal insight into the psychopathology of fear-based disorders.

A limitation of this study is that we only had explicit fear learning as measured by verbal self-report contingency awareness data. Because this was an exploratory analysis arising from an unexpected finding from an overarching study, we did not specifically model explicit and implicit learning as a variable of interest within the original study design. Skin conductance response could have been used as an indicator of implicit learning, and SCR is a common method for measuring conditioned fear responses. We did collect SCR data; however, our SCR data were not viable for analysis due to technological errors with sampling, which are not uncommon when collecting SCR data in the scanner environment (Bjorkstrand, 1990; Indovina et al., 2011; Pohlack, Nees, Liebscher, et al., 2012; Pohlack, Nees, Ruttorf, et al., 2012). Therefore, we did not model for and cannot draw conclusion about implicit learning in the present study, nor contrast between regions and networks, which may be involved in implicit versus explicit learning. Usable simultaneously collected SCR data also would have provided some evidence as to whether participants were engaged in the task. With this lack of behavioural data and having only collected expectancy data after conditioning as a single measure instead of throughout the task, there is potential for the results presented herein to reflect a general lack of engagement in the task (i.e., no learning) as opposed to a lack of explicit learning. However, in the unaware group, expectancies of the US were on the average or higher end of the scale for *both* CSs, rather than being low—it appears that the difference is a lack of differentiation between the CS+ and the CS− in the unaware group compared with the aware group. There was a larger response to the CS− in the unaware group compared with the aware group. In other words, such lack of difference could be a result of lack of discrimination, i.e., overgeneralization of learned fear. The small sample assessed in this study is another limitation, and as such we were underpowered to conduct any sex-specific analyses even though there are known sex-related differences in fear learning. Assessing a potential moderating role of sex on contingency awareness in fear learning could be a new route of exploration. This small sample also limits the generalizability of our findings—particularly because our sample was a young, college-educated cohort. Finally, we did not explore differences in contingency awareness in any clinical population.

To our knowledge, this is one of few functional neuroimaging studies that extensively looked at the neurocircuitry of contingency awareness in Pavlovian fear conditioning and the first to examine network connectivity. We found involvement of regions composing the salience network involved in threat detection and fear leaning and the CEN involved in working memory, attention, and task completion. Our findings confirm prior work indicating the hippocampus may have a pivotal role in the underlying neurocircuitry of contingency awareness of Pavlovian fear conditioning. Future studies in patient populations will be significant in understanding explicit fear learning in fear-based psychopathologies.

Declarations

References

Adamson, L., Hartman, S. G., & Lyxell, B. (1999). Adolescent identity-a qualitative approach: Self-concept, existential questions and adult contacts. *Scandinavian Journal of Psychology*, *40*(1), 21-31. https://doi.org/10.1111/1467-9450.00094

Adolphs, R. (2013). The biology of fear. *Curr Biol*, *23*(2), R79-93. https://doi.org/10.1016/j.cub.2012.11.055

Baeuchl, C., Hoppstädter, M., Meyer, P., & Flor, H. (2019). Contingency awareness as a prerequisite for differential contextual fear conditioning. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 811-828. https://doi.org/10.3758/s13415-018-00666-z

Balderston, N. L., & Helmstetter, F. J. (2010). Conditioning with masked stimuli affects the timecourse of skin conductance responses. *Behavior Neuroscience*, *124*(4), 478-489. https://doi.org/10.1037/a0019927

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*(5227), 1115-1118. https://doi.org/10.1126/science.7652558

Bjorkstrand, P. A. (1990). Effects of conditioned stimulus pre-exposure on human electrodermal conditioning to fear-relevant and fear-irrelevant stimuli. *Biology Psychology*, *30*(1), 35-50. https://doi.org/10.1016/0301-0511(90)90089-f

Briscione, M., Jovanovic, T., & Norrholm, S. (2014). Conditioned Fear Associated Phenotypes as Robust, Translational Indices of Trauma-, Stressor-, and Anxiety-Related Behaviors. *Frontiers in Psychiatry*, *5*, 88. https://doi.org/10.3389/fpsyt.2014.00088

Carter, R. M., Hofstotter, C., Tsuchiya, N., & Koch, C. (2003). Working memory and fear conditioning. *Process National Academic Science USA*, *100*(3), 1399-1404. https://doi.org/10.1073/pnas.0334049100

Carter, R. M., O'Doherty, J. P., Seymour, B., Koch, C., & Dolan, R. J. (2006). Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *NeuroImage*, *29*(3), 1007-1012. https://doi.org/10.1016/j.neuroimage.2005.09.011

Casey, B. J., Cohen, J. D., Jezzard, P., Turner, R., Noll, D. C., Trainor, R. J., Giedd, J., Kaysen, D., Hertz-Pannier, L., & Rapoport, J. L. (1995). Activation of Prefrontal Cortex in Children during a Nonspatial Working Memory Task with Functional MRI. *NeuroImage*, *2*(3), 221-229. https://doi.org/10.1006/nimg.1995.1029

Cheng, D. T., Knight, D. C., Smith, C. N., Stein, E. A., & Helmstetter, F. J. (2003). Functional MRI of human amygdala activity during Pavlovian fear conditioning: stimulus processing versus response expression. *Behavior Neuroscience*, *117*(1), 3-10. https://doi.org/10.1037//0735-7044.117.1.3

Cocchi, L., Bramati, I. E., Zalesky, A., Furukawa, E., Fontenelle, L. F., Moll, J., Tripp, G., & Mattos, P. (2012). Altered functional brain connectivity in a non-clinical sample of young adults with attention-deficit/hyperactivity disorder. *Journal of Neuroscience*, *32*(49), 17753-17761. https://doi.org/10.1523/JNEUROSCI.3272-12.2012

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron*, *33*(4), 653-663. https://doi.org/10.1016/s0896-6273(02)00588-3

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cognition Science*, *7*(9), 415-423. https://www.ncbi.nlm.nih.gov/pubmed/12963473

Cushman, J. D., & Fanselow, M. S. (2010). Fear Conditioning. In G. F. Koob, M. L. Moal, & R. F. Thompson (Eds.), *Encyclopedia of Behavioral Neuroscience* (pp. 524-531). Academic Press. https://doi.org/10.1016/B978-0-08-045396-5.00136-6

Davis, M. (1997). Neurobiology of fear responses: the role of the amygdala. *Journal of Neuropsychiatry Clinic Neuroscience*, *9*(3), 382-402. https://doi.org/10.1176/jnp.9.3.382

Dere, E., Pause, B. M., & Pietrowsky, R. (2010). Emotion and episodic memory in neuropsychiatric disorders. *Behav Brain Res*, *215*(2), 162-171. https://doi.org/10.1016/j.bbr.2010.03.017

Evans, G. W., Swain, J. E., King, A. P., Wang, X., Javanbakht, A., Ho, S. S., Angstadt, M., Phan, K. L., Xie, H., & Liberzon, I. (2016). Childhood cumulative risk exposure and adult amygdala volume and function. *Journal of Neuroscience Research*, *94*(6), 535-543.

Friedman, A. L., Burgess, A., Ramaseshan, K., Easter, P., Khatib, D., Chowdury, A., Arnold, P. D., Hanna, G. L., Rosenberg, D. R., & Diwadkar, V. A. (2017). Brain network dysfunction in youth with obsessive-compulsive disorder induced by simple uni-manual behavior: The role of the dorsal anterior cingulate cortex. *Psychiatry Res Neuroimaging*, *260*, 6-15. https://doi.org/10.1016/j.pscychresns.2016.12.005

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage*, *6*(3), 218-229. https://doi.org/10.1006/nimg.1997.0291

Goodman, A. M., Harnett, N. G., & Knight, D. C. (2018a). Pavlovian conditioned diminution of the neurobehavioral response to threat. *Neurosci Biobehav Rev*, *84*, 218-224. https://doi.org/10.1016/j.neubiorev.2017.11.021

Goodman, A. M., Harnett, N. G., Wheelock, M. D., Hurst, D. R., Orem, T. R., Gossett, E. W., Dunaway, C. A., Mrug, S., & Knight, D. C. (2018b). Anticipatory prefrontal cortex activity underlies stress-induced changes in Pavlovian fear conditioning. *NeuroImage*, *174*, 237-247. https://doi.org/10.1016/j.neuroimage.2018.03.030

Greco, J. A., & Liberzon, I. (2016). Neuroimaging of Fear-Associated Learning. *Neuropsychopharmacology*, *41*(1), 320-334. https://doi.org/10.1038/npp.2015.255

Gu, X., Hof, P. R., Friston, K. J., & Fan, J. (2013). Anterior insular cortex and emotional awareness. *J Comp Neurol*, *521*(15), 3371-3388. https://doi.org/10.1002/cne.23368

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *Int J Psychophysiol*, *57*(1), 5-14. https://doi.org/10.1016/j.ijpsycho.2005.01.006

Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: implications for exposure therapy of anxiety disorders. *Clinical Psychology Review*, *28*(2), 199-210. https://doi.org/10.1016/j.cpr.2007.04.009

Indovina, I., Robbins, T. W., Nunez-Elizalde, A. O., Dunn, B. D., & Bishop, S. J. (2011). Fear-conditioning mechanisms associated with trait vulnerability to anxiety in humans. *Neuron*, *69*(3), 563-571. https://doi.org/10.1016/j.neuron.2010.12.034

Javanbakht, A., Duval, E. R., Cisneros, M. E., Taylor, S. F., Kessler, D., & Liberzon, I. (2017). Instructed fear learning, extinction, and recall: additive effects of cognitive information on emotional learning of fear. *Cognition and Emotion*, *31*(5), 980-987. https://doi.org/10.1080/02699931.2016.1169997

Javanbakht, A., Grasser, L. R., Madaboosi, S., Chowdury, A., Liberzon, I., & Diwadkar, V. A. (2021). The Neurocircuitry Underlying Additive Effects of Safety Instruction on Extinction Learning [10.3389/fnbeh.2020.576247]. *Frontiers in Behavioral Neuroscience*, *14*, 263. https://www.frontiersin.org/article/10.3389/fnbeh.2020.576247

Javanbakht, A., Kim, P., Swain, J. E., Evans, G. W., Phan, K. L., & Liberzon, I. (2016). Sex-specific effects of childhood poverty on neurocircuitry of processing of emotional cues: a neuroimaging study. *Behavioral Sciences*, *6*(4), 28.

Javanbakht, A., King, A. P., Evans, G. W., Swain, J. E., Angstadt, M., Phan, K. L., & Liberzon, I. (2015). Childhood poverty predicts adult amygdala and frontal activity and connectivity in response to emotional faces. *Frontiers in Behavioral Neuroscience*, *9*, 154.

Kattoor, J., Gizewski, E. R., Kotsis, V., Benson, S., Gramsch, C., Theysohn, N., Maderwald, S., Forsting, M., Schedlowski, M., & Elsenbruch, S. (2013). Fear Conditioning in an Abdominal Pain Model: Neural Responses during Associative Learning and Extinction in Healthy Subjects. *PLOS ONE*, *8*(2), e51149. https://doi.org/10.1371/journal.pone.0051149

Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., & Stark, R. (2009a). Neural, electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture-picture conditioning paradigm. *Neuroscience*, *158*(2), 721-731. https://doi.org/10.1016/j.neuroscience.2008.09.049

Klucken, T., Tabbert, K., Schweckendiek, J., Merz, C. J., Kagerer, S., Vaitl, D., & Stark, R. (2009b). Contingency learning in human fear conditioning involves the ventral striatum. *Hum Brain Mapp*, *30*(11), 3636-3644. https://doi.org/10.1002/hbm.20791

Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2003). Expression of conditional fear with and without awareness. *Proc Natl Acad Sci U S A*, *100*(25), 15280-15283. https://doi.org/10.1073/pnas.2535780100

Knight, D. C., Waters, N. S., & Bandettini, P. A. (2009). Neural substrates of explicit and implicit fear memory. *NeuroImage*, *45*(1), 208-214. https://doi.org/10.1016/j.neuroimage.2008.11.015

Knight, D. C., & Wood, K. H. (2011). Investigating the neural mechanisms of aware and unaware fear memory with FMRI. *J Vis Exp*(56), 1-6. https://doi.org/10.3791/3083

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998-1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Kropotov, J. D. (2009). Chapter 13 - Affective System. In J. D. Kropotov (Ed.), *Quantitative EEG, Event-Related Potentials and Neurotherapy* (pp. 292-309). Academic Press. https://doi.org/10.1016/B978-0-12-374512-5.00013-X

LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *J Neurosci*, *15*(10), 6846-6855. https://doi.org/10.1523/JNEUROSCI.15-10-06846.1995

Labrenz, F., Icenhour, A., Benson, S., & Elsenbruch, S. (2015). Contingency Awareness Shapes Acquisition and Extinction of Emotional Responses in a Conditioning Model of Pain-Related Fear. *Front Behav Neurosci*, *9*, 318. https://doi.org/10.3389/fnbeh.2015.00318

Ladouceur, C. D., Farchione, T., Diwadkar, V., Pruitt, P., Radwan, J., Axelson, D. A., Birmaher, B., & Phillips, M. L. (2011). Differential Patterns of Abnormal Activity and Connectivity in the Amygdala–Prefrontal Circuitry in Bipolar-I and Bipolar-NOS Youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(12), 1275-1289.e1272. https://doi.org/10.1016/j.jaac.2011.09.023

Leuchs, L., Schneider, M., & Spoormaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, *56*(1), e13283. https://doi.org/10.1111/psyp.13283

Lévesque, J., Eugène, F., Joanette, Y., Paquette, V., Mensour, B., Beaudoin, G., Leroux, J.-M., Bourgouin, P., & Beauregard, M. (2003). Neural circuitry underlying voluntary suppression of sadness. *Biological Psychiatry*, *53*(6), 502-510. https://doi.org/10.1016/S0006-3223(02)01817-6

Liberzon, I., & Abelson, J. L. (2016). Context Processing and the Neurobiology of Post-Traumatic Stress Disorder. *Neuron*, *92*(1), 14-30. https://doi.org/10.1016/j.neuron.2016.09.039

Liberzon, I., & Ressler, K. (2012). *Generalization of Learned Fear from The Neurobiology of PTSD*. The National Academies Press. https://doi.org/10.17226/13364

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *The Behavioral and brain sciences*, *35*(3), 121-143. https://doi.org/10.1017/S0140525X11000446

Lovibond, P. F. (2019). Cognitive Processes in Extinction. *Learn Mem*, *26*(8).

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal Experimental Psychology Animal Behavior Process*, *28*(1), 3-26. https://www.ncbi.nlm.nih.gov/pubmed/11868231

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*

Marin, M.-F., Zsido, R. G., Song, H., Lasko, N. B., Killgore, W. D. S., Rauch, S. L., Simon, N. M., & Milad, M. R. (2017). Skin Conductance Responses and Neural Activations During Fear Conditioning and Extinction Recall Across Anxiety Disorders. *JAMA Psychiatry*, *74*(6), 622-631. https://doi.org/10.1001/jamapsychiatry.2017.0329

MathWorks (2013;2019). *MATLAB* (Version R2013b, R2019a) [Numerical computing]. https://www.mathworks.com/

McNamara, R. K., Liu, Y., Jandacek, R., Rider, T., & Tso, P. (2008). The aging human orbitofrontal cortex: Decreasing polyunsaturated fatty acid composition and associated increases in lipogenic gene expression and stearoyl-CoA desaturase activity. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, *78*(4), 293-304. https://doi.org/10.1016/j.plefa.2008.04.001

Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. *Neuroimage*, *49*(2), 1760-1768. https://doi.org/10.1016/j.neuroimage.2009.09.040

Milad, M. R., Orr, S. P., Lasko, N. B., Chang, Y., Rauch, S. L., & Pitman, R. K. (2008). Presence and acquired origin of reduced recall for fear extinction in PTSD: results of a twin study. *Journal Psychiatry Research*, *42*(7), 515-520. https://doi.org/10.1016/j.jpsychires.2008.01.017

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, *42*(4), 456-464. https://doi.org/10.1111/j.1469-8986.2005.00302.x

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., Zeidan, M. A., Handwerger, K., Orr, S. P., & Rauch, S. L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biology Psychiatry*, *66*(12), 1075-1082. https://doi.org/10.1016/j.biopsych.2009.06.026

Mylius, V., Ayache, S. S., Ahdab, R., Farhat, W. H., Zouari, H. G., Belke, M., Brugières, P., Wehrmann, E., Krakow, K., Timmesfeld, N., Schmidt, S., Oertel, W. H., Knake, S., & Lefaucheur, J. P. (2013). Definition of DLPFC and M1 according to anatomical landmarks for navigated brain stimulation: Inter-rater reliability, accuracy, and influence of gender and age. *Neuroimage*, *78*, 224-232. https://doi.org/10.1016/j.neuroimage.2013.03.061

Neuroimaging, W. C. f. H (2020). *SPM12*. Functional Imaging Laboratory. https://www.fil.ion.ucl.ac.uk/spm/

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604-609. https://doi.org/10.1093/scan/nss055

O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends in Neurosciences*, *33*(8), 355-361. https://doi.org/10.1016/j.tins.2010.05.002

Pellman, B. A., & Kim, J. J. (2015). Fear: Psychological and Neural Aspects. *International Encyclopedia of the Social & Behavioral Sciences:* , 868-874. https://doi.org/10.1016/B978-0-08-097086-8.55024-7

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198-202. https://doi.org/10.1016/j.conb.2004.03.015

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction Learning in Humans: Role of the Amygdala and vmPFC. *Neuron*, *43*(6), 897-905. https://doi.org/10.1016/j.neuron.2004.08.042

Picard, F. (2013). State of belief, subjective certainty and bliss as a product of cortical dysfunction. *Cortex*, *49*(9), 2494-2500. https://doi.org/10.1016/j.cortex.2013.01.006

Pohlack, S. T., Nees, F., Ruttorf, M., Schad, L. R., & Flor, H. (2012b). Activation of the ventral striatum during aversive contextual conditioning in humans. *Biology Psychology*, *91*(1), 74-80. https://doi.org/10.1016/j.biopsycho.2012.04.004

Qiu, Y.-W., Lv, X.-F., Jiang, G.-H., Su, H.-H., Yu, T., Tian, J.-Z., Zhang, X.-L., & Zhuo, F.-Z. (2014). Reduced ventral medial prefrontal cortex (vmPFC) volume and impaired vmPFC-default mode network integration in codeine-containing cough syrups users. *Drug and Alcohol Dependence*, *134*, 314-321. https://doi.org/10.1016/j.drugalcdep.2013.10.023

Rabinak, C. A., Mori, S., Lyons, M., Milad, M. R., & Phan, K. L. (2017). Acquisition of CS-US contingencies during Pavlovian fear conditioning and extinction in social anxiety disorder and posttraumatic stress disorder. *Journal of Affect Disorders*, *207*, 76-85. https://doi.org/10.1016/j.jad.2016.09.018

Raij, T., Nummenmaa, A., Marin, M.-F., Porter, D., Furtak, S., Setsompop, K., & Milad, M. R. (2018). Prefrontal Cortex Stimulation Enhances Fear Extinction Memory in Humans. *Biological Psychiatry*, *84*(2), 129-137. https://doi.org/10.1016/j.biopsych.2017.10.022

Rilling, J. K., & Sanfey, A. G. (2009). Social Interaction. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 41-48). Academic Press. https://doi.org/10.1016/B978-008045046-9.01539-4

Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M. (2018). Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency Activity in Human OFC. *Current Biology*, *28*(18), 2889-2899.e2883. https://doi.org/10.1016/j.cub.2018.07.045

Sanders, M. J., Wiltgen, B. J., & Fanselow, M. S. (2003). The place of the hippocampus in fear conditioning. *European Journal of Pharmacology*, *463*(1), 217-223. https://doi.org/10.1016/S0014-2999(03)01283-4

Schultz, D. H., Balderston, N. L., Geiger, J. A., & Helmstetter, F. J. (2013). Dissociation between implicit and explicit responses in postconditioning UCS revaluation after fear conditioning in humans. *Behavior Neuroscience*, *127*(3), 357-368. https://doi.org/10.1037/a0032742

Selemon, L. D., Young, K. A., Cruz, D. A., & Williamson, D. E. (2019). Frontal Lobe Circuitry in Posttraumatic Stress Disorder. *Chronic Stress (Thousand Oaks)*, *3*. https://doi.org/10.1177/2470547019850166

Sevenster, D., Beckers, T., & Kindt, M. (2014). Fear conditioning of SCR but not the startle reflex requires conscious discrimination of threat and safety. *Front Behavior Neuroscience*, *8*, 32. https://doi.org/10.3389/fnbeh.2014.00032

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(Suppl 20), 22-33.

Silvers, J. A., Wager, T. D., Weber, J., & Ochsner, K. N. (2015). The neural bases of uninstructed negative emotion modulation. *Society Cognitive Affect Neuroscience*, *10*(1), 10-18. https://doi.org/10.1093/scan/nsu016

Silverstein, B. H., Bressler, S. L., & Diwadkar, V. A. (2016). Inferring the Dysconnection Syndrome in Schizophrenia: Interpretational Considerations on Methods for the Network Analyses of fMRI Data. *Frontiers in Psychiatry*, *7*. https://doi.org/10.3389/fpsyt.2016.00132

Sperl, M. F. J., Panitz, C., Hermann, C., & Mueller, E. M. (2016). A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials. *Psychophysiology*, *53*(9), 1352-1365. https://doi.org/10.1111/psyp.12677

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195-231. https://doi.org/10.1037/0033-295X.99.2.195

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, *105*(34), 12569-12574. https://doi.org/10.1073/pnas.0800005105

Staudinger, M. R., Erk, S., & Walter, H. (2011). Dorsolateral prefrontal cortex modulates striatal reward encoding during reappraisal of reward anticipation. *Cereb Cortex*, *21*(11), 2578-2588. https://doi.org/10.1093/cercor/bhr041

Sturm, V. E., Haase, C. M., & Levenson, R. W. (2016). Emotional Dysfunction in Psychopathology and Neuropathology: Neural and Genetic Pathways. *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*, 345-364. https://doi.org/10.1016/B978-0-12-800105-9.00022-6

Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T., & Stark, R. (2011). Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. *Society Cognitive Affective Neuroscience*, *6*(4), 495-506. https://doi.org/10.1093/scan/nsq070

Tabbert, K., Stark, R., Kirsch, P., & Vaitl, D. (2006). Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. *NeuroImage*, *32*(2), 761-770. https://doi.org/10.1016/j.neuroimage.2006.03.038

Thomson, P., & Jaque, S. V. (2017). 6 - Neurobiology, creativity, and performing artists. In P. Thomson & S. V. Jaque (Eds.), *Creativity

and the Performing Artist (pp. 79-102). Academic Press. https://doi.org/10.1016/B978-0-12-804051-5.00006-8

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15*(1), 273-289. https://doi.org/10.1006/nimg.2001.0978

Uddin, L. Q. (2016). *Salience network of the human brain* (1 ed.). Academic Press is an imprint of Elsevier.

Ward, B. (2000). *Simultaneous Inference for fMRI Data*.

Watanabe, M. (2017). Brodmann Areas 8 and 9 Including the Frontal Eye Field *Brain Nerve, 69*(4), 347-354. https://doi.org/10.11477/mf.1416200751

Weike, A. I., Schupp, H. T., & Hamm, A. O. (2007). Fear acquisition requires awareness in trace but not delay conditioning. *Psychophysiology, 44*(1), 170-180. https://doi.org/10.1111/j.1469-8986.2006.00469.x

Yin, S., Liu, Y., Petro, N. M., Keil, A., & Ding, M. (2018). Amygdala Adaptation and Temporal Dynamics of the Salience Network in Conditioned Fear: A Single-Trial fMRI Study. *eNeuro, 5*(1). https://doi.org/10.1523/ENEURO.0445-17.2018