



Original article

Implicit threat learning involves the dorsolateral prefrontal cortex and the cerebellum

Charlene L.M. Lam^{a,b,*,#}, Clive H.Y. Wong^{b,c,f,*}, Markus Junghöfer^d, Kati Roesmann^{d,e}^a Laboratory of Clinical Psychology and Affective Neuroscience, Department of Psychology, The University of Hong Kong, Hong Kong, China^b State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China^c Laboratory of Neuropsychology and Human Neuroscience, The University of Hong Kong, Hong Kong, China^d Institute for Biomagnetism and Biosignalanalysis, University of Münster, Germany^e Institute of Clinical Psychology and Psychotherapy, University of Siegen, Germany^f Department of Psychology, The Education University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:

Implicit, fear conditioning, cerebellum, contingency awareness
Multi-cs conditioning

ABSTRACT

Background/Objective: Most studies investigating the neural correlates of threat learning were carried out using an explicit Pavlovian conditioning paradigm where declarative knowledge on contingencies between conditioned (CS) and unconditioned stimuli (US) is acquired. The current study aimed at understanding the neural correlates of threat conditioning when contingency awareness is limited or even absent.**Method:** We conducted an fMRI report of threat learning in an implicit associative learning paradigm called multi-CS conditioning, in which a number of faces were associated with aversive screams (US) such that participants could not report contingencies between the faces and the screams.**Results:** The univariate results showed support for the recruitment of threat-related regions including the dorsolateral prefrontal cortex (dlPFC) and the cerebellum during acquisition. Further analyses by the multivariate representational similarity technique identified learning-dependent changes in the bilateral dlPFC.**Conclusion:** Our findings support the involvement of the dlPFC and the cerebellum in threat conditioning that occurs with highly limited or even absent contingency awareness.

Learning to predict danger and forming associative memories are crucial and adaptive for survival. Given its evolutionary importance, researchers have long studied the neurobiological processes underlying the ability to associate environmental stimuli with an aversive outcome. Typically, such associative learning processes are modelled in fear conditioning paradigms where an initially neutral stimulus (conditioned stimulus, CS) come to elicit a conditioned response after being paired with an intrinsically aversive unconditioned stimulus (US). One significant insight drawn from fear conditioning studies is that the neural network of fear-associated learning is relatively well conserved across species (Greco & Liberzon, 2016; Maren & Holmes, 2016; Maren, Phan & Liberzon, 2013).

A detailed account of the fear conditioning neural network has been extensively studied (Feinstein, Adolphs, Damasio & Tranel, 2011; Ledoux & Daw, 2018). In brief, the amygdala and the hippocampus are implicated in the learning and expression of fear (Phelps, 2006). The bidirectional communication between the two regions is integral in the

encoding and processing of the contexts associated with fear (Fanselow, 2000; Sparta et al., 2014) and the acquisition of the conditioned responses (Andreatta et al., 2015; Bach, Weiskopf & Dolan, 2011). In addition, the prefrontal cortex (PFC) – the ventral, medial, and dorsolateral subregions – are involved and play an important role in threat-safety discrimination, the anticipation of threat, as well as the subjective appraisal of fear and anxiety (Kalisch & Gerlicher, 2014; Lau & Rosenthal, 2011; Milad & Quirk, 2012).

Recent fMRI studies using Pavlovian conditioning as a model of fear-related associative learning have supported the role of the cerebellum in threat learning and processing (Batsikadze et al., 2022; Ernst et al., 2019; Kattoor et al., 2014). Several recent studies reported that the cerebellum is highly topographically arranged, i.e. different parts of the cerebellum are involved in different aspects of motor and non-motor functions (King, Hernandez-Castillo, Poldrack, Ivry & Diedrichsen, 2019; Pierce & Péron, 2020; Stoodley & Schmahmann, 2018; Strata, 2015). In light of threat learning, the midline parts of the cerebellum

Corresponding author: Charlene L.M. Lam, Ph.D., Department of Psychology, The University of Hong Kong, Rm 6.59, The Jockey Club Tower, Pokfulam Road, Hong Kong, China Tel.: (852) 3917-2385.

E-mail address: charlene.lam@hku.hk (C.L.M. Lam).

* These authors share first authorship

<https://doi.org/10.1016/j.ijchp.2022.100357>

Received 15 September 2022; Accepted 14 November 2022

Available online 23 November 2022

1697-2600/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

such as the vermis are likely involved in autonomic processes, and the posterolateral parts such as lobules Crus I and VI in cognitive and higher-order emotional processes and prediction of harmful stimuli (Ernst et al., 2019).

Contingency awareness refers to the phenomenon of becoming consciously aware of the association between a CS and an US. To facilitate a translational approach, most previous research on fear conditioning has focused on relatively simple learning paradigms, in which only one stimulus is paired with an US in the conditioning process. Yet, a simple single CS to single US associative learning situation in real life is rare. For example, in the scene of a traumatic car accident, multiple CS (e.g. the traffic lights, the inflated airbag) are often capable to elicit conditioned fear responses. On an experimental level, such scenario could be modelled in a multi-CS conditioning paradigm (Steinberg et al., 2012).

Multi-CS conditioning pairs a multitude of neutral stimuli with one or multiple unconditioned stimuli during acquisition, forming multiple conditioned stimuli. Because of the large number of stimuli involved in learning, this paradigm allows the investigation of the implicit processes in affective learning under conditions of limited CS-US awareness. Previous multi-CS studies have yielded successful threat acquisition and extinction (Brockelmann et al., 2011; Roesmann et al., 2020; Steinberg et al., 2012). On neural levels, CS+ (i.e. US associated CS) evoked stronger magnetoencephalographic responses in lateral and orbital prefrontal regions, as well as in temporo-occipital (Steinberg, Bröckelmann, Rehbein, Döbel & Junghöfer, 2013). The involvement of subcortical and cerebellar regions and their interplay with the prefrontal cortex in this implicit form of fear learning has remained unclear – mainly due to the limited depth and spatial resolution of the MEG for activity in these regions.

In this fMRI study, we investigated the neural correlates of threat associations that occur under very limited contingency awareness using the multi-CS conditioning paradigm. To this end, we presented 54 different neutral facial stimuli which were either paired with aversive screams or left unpaired in the scanner and detected the changes of BOLD signals within the fear learning circuitry. Based on previous studies showing the involvement of the prefrontal cortex, the amygdala, the hippocampus and the cerebellum in explicit fear conditioning (Ernst et al., 2019; Fullana et al., 2018; Milad & Quirk, 2012; Steinberg et al., 2013), we hypothesised that higher activations in the cortical (PFC), subcortical (amygdala, hippocampus), and the cerebellar regions would be observed in the CS+ relative to the CS- (i.e. CS never paired with US) during the acquisition of multi-CS conditioning. We applied univariate analysis to answer whether multi-CS conditioning, being a relatively novel design to study threat learning under very limited or an absence of contingency awareness, would activate similar threat conditioning regions as those found in traditional single-CS conditioning paradigms. Given the high number of pairings in the multi-CS conditioning paradigm in the acquisition phase, we applied multivariate representational similarity analyses (RSA) to further assess how multi-CS/US associations emerged as a function of repeated pairings. We hypothesised that the multivoxel pattern in areas associated with threat learning would show a converging similarity against time with the acquired threat association pattern.

Methods

Participants

Twenty-two healthy university students (25.68 ± 3.93 years; males: females = 11:11) were recruited in this study. All participants reported normal hearing, normal or corrected-to-normal vision. They were excluded if they self-reported current or history of psychiatric/neurological illnesses, and showed contraindications for undergoing MR imaging: 1) installation of metallic implants or medical apparatus (e.g. pacemakers or artificial joints) in the body, 2) claustrophobia, 3) pregnancy or 4) having tattoos. No participants were excluded because of conspicuous scores on the Positive and Negative Affect Scale (PANAS; Watson et

al., 1988) and State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983) in the current study. The study protocol was approved by the ethics committee of the XXXX and adhered to the tenets of the Declaration of Helsinki. Written informed consent forms were obtained from all participants.

Materials

Unconditioned stimuli

Two aversive sounds (duration: 1000 ms; USA: a female scream, USB: a male scream) were purchased from the company Pond5 (Pond5 Media Ireland Ltd.). The screams were normalised and resampled to 44,100 Hz. Human screams were used in previous multi-CS conditioning paradigms (e.g. (Roesmann et al., 2020)) and were chosen in our experiment as well because they had relevant associations with the CSs in our experiment, which were all female or male faces. They were delivered binaurally through MRI-compatible headphones inside the scanner. The assignment of the US was balanced across participants.

Conditioned stimuli

Fifty-four images displaying faces (27 females) with neutral expressions were selected from existing face databases, including the Karolinska Directed Emotional Faces archive (Lundqvist, Flykt & Öhman, 1998) and the NimStim Face Stimulus Set (Tottenham et al., 2009). All faces were converted to grayscale images in Adobe® Photoshop® and were adjusted for their levels of brightness and contrast. They were pseudo-randomly split into three conditions: 18 CS+ (CSa+, paired with USA), 18 CS+ (CSb+, paired with USB) and 18 CS- faces (unpaired during conditioning). The attractiveness of the faces was rated by a pilot group of participants, who were recruited separately ($N = 20$); there were no significant differences in the rated attractiveness of the faces across the three sets of faces that were counterbalanced in three conditions ($F(2, 57) = 1.00, p = .376$).

CS-US matching task

Explicit knowledge of the stimulus category was assessed using a computerised CS-US matching task after Acquisition (Figure S1). All 54 CS were pseudo-randomly presented for 600 ms, followed by two questions. First, participants were asked to indicate for each face whether it was paired with a scream during conditioning (stimulus category: CS+ vs. CS-) on a Likert scale from -4 (surely there was no scream) to 4 (surely there was a scream). Higher scores indicated higher confidence about their response. Second, they were asked to indicate by forced choice (also for the assumed CS-) whether the faces were paired with a male scream or a female scream on a Likert scale (-4 = surely female to 4 = surely male). Again, higher absolute scores represented higher confidence in their response. For practice, they completed three trials with later unused faces prior to the start of the task.

Pair comparison task

This preference rating task served as an indirect measure to evaluate participants' awareness of CS-US contingency pairings after Acquisition (Figure S1). On this task, pairs of CS+ and CS- faces were shown on the computer screen. For each pair of same-gender faces, participants were asked to decide which face they preferred in a binary forced-choice format, in which they indicated their preference by ticking the box beneath the corresponding face. No time limit was set on this task but participants were encouraged to respond as quickly as they could. Participants responded to a total of 54 comparisons (27 comparisons for female faces and 27 comparisons for male faces) after Acquisition. Participants endorsing the CS- faces might suggest a relative positive hedonic valence of the faces, hence a relative awareness of the CS-US contingency.

The conditioning stimuli were presented in the MRI scanner using E-Prime (Psychology Software Tools, Pittsburgh, PA). CS-US matching

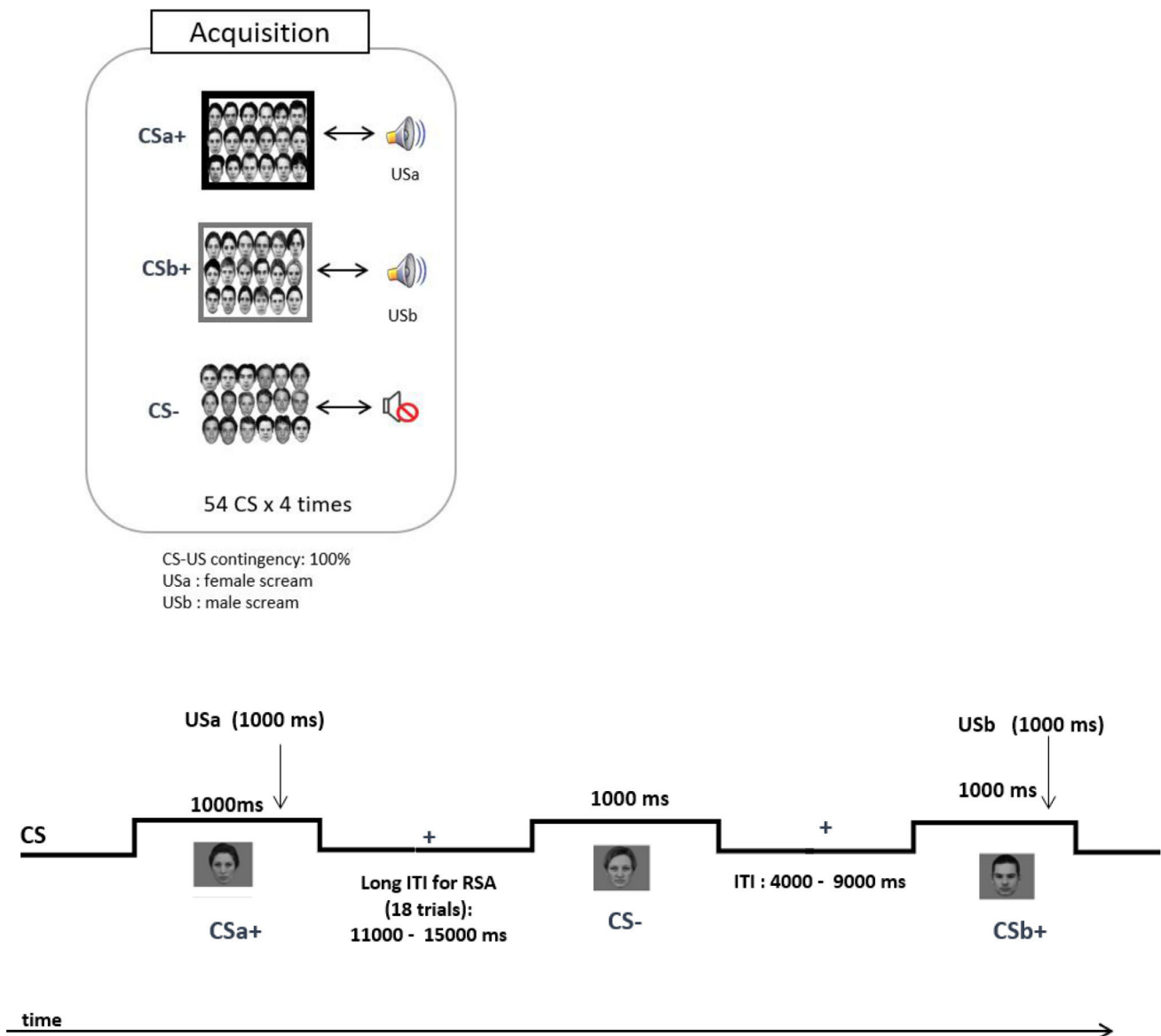


Fig. 1. Experimental timeline during acquisition in the scanner. A total of 54 grayscale images displaying faces (27 females) with neutral expressions were pseudo-randomly split into three conditions: 18 CSa+ (paired with a female scream: USa), 18 CSb+ (paired with a male scream: USb) and 18 CS- faces (unpaired during conditioning). Each CS was presented 4 times ($54 \times 4 = 216$ trials). All CS were presented for 1000 ms. The onset of the US was jittered from 200 to 700 ms following the onset of the CS which lasted for 1000 ms. The inter-stimulus interval (ITI) varied pseudo-randomly between 4000 and 15,000 ms. CS: conditioned stimulus; US: unconditioned stimulus; RSA: Representational Similarity Analysis.

task and Pair comparison task were conducted on a computer using Presentation (Neurobehavioural Systems, Albany, CA).

Design and procedure

The data analysed in this report was taken from a larger project investigating the reconsolidation of fear memories in a multi-CS conditioning paradigm. The experiment consisted of three sessions that took place on three successive days: Acquisition (Day 1), Reactivation and Extinction (Day2), and Reinstatement and Re-extinction (Day 3). The total experimental time was about 5 h. The current study focuses on the Acquisition phase (Day 1) because the multi-CS conditioning paradigm is a relatively novel paradigm for studying implicit threat learning. The role of subcortical circuits during the acquisition of multiple, largely implicit CS-US associations has remained unknown, mainly due to the lack of fMRI studies with a good spatial resolution in deep structures. Details of the acquisition phase are described below and in Fig. 1.

Acquisition

Participants first completed the questionnaires (see supplementary materials) and habituation of the CS outside the scanner, followed by acquisition inside the scanner. The habituation mainly served to familiarize participants with the procedure of the experiment. Each CS was presented for 1000 ms, with an inter-trial interval of 4000 to 15,000 ms (Fig. 1). The CS+ -US contingency was maintained at 100% and the CS- was always presented without the US. We jittered the onset of the US from 200 to 700 ms following the onset of the CSs. The presentation order, the time of onset of the CS, and the time of onset of the US were arranged with reference to the functional MRI of the Brain (FMRIB) Software Library (FSL)'s design efficiency (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012), in which 5000 sequences were produced, and the model with the highest efficiency for estimation of the BOLD response was chosen. During acquisition, each CS was presented four times (216 trials in total), and the overall presentations were divided into four runs (8.1 to 8.7 min each, 33.9 mins in total) (Fig. 1). In each

run, 18 (12 CS+, 6 CS-) out of 54 trials had extended interstimulus intervals (9 s to 15 s) to reduce intrinsic noise correlations. These trials were designed for running the RSA to capture the process of learning. The following general instruction was given to the participants: "Some faces will be followed by a female scream; some will be followed by a male scream, and some will not be followed by any scream. You will be able to figure out which ones by paying close attention." The assignment of the set of faces to each CS condition was counterbalanced across participants. The imaging parameters are specified below:

Brain imaging

Data were acquired using a 3T Philips Achieva MR scanner (Philips Medical Systems, The Netherlands) equipped with an 8-channel SENSE head coil. Head movement was restricted using foam cushions. High-resolution anatomical T1-weighted images were acquired using a magnetization-prepared rapid gradient echo (MPRAGE) sequence with the following parameters: 160 sagittal slices, repetition time (TR) = 6.9 ms, echo time (TE) = 3.2 ms, matrix = 240×240 , FOV = $240 \times 240 \times 160$ mm³, flip angle = 8°, voxel size = $1 \times 1 \times 1$ mm³. During visual presentations, task-based BOLD imaging was collected using a T2*-weighted echo-planar imaging (EPI) sequence (39 slices, TE = 30 ms, TR = 2000 ms, matrix = 124×124 , FOV = 230×230 mm², flip angle = 90°, voxel size = $1.6 \times 1.6 \times 3.5$ mm³). Each CS was presented four times (216 trials in total), and the overall presentation was divided into four runs (8.1 to 8.7 min each). The duration of the scans was 33.9 min during acquisition.

Following the acquisition, participants completed the Pair Comparison Task and CS-US Matching Task outside the scanner.

Statistical analysis

Contingency awareness: behavioural tasks

For the CS-US matching task, the sensitivity index d' (Green & Swets, 1966) was employed to detect how well participants recognised the stimulus category of each CS. First, frequencies of correct matches were calculated to obtain the hits and false alarm rates, as well as the d' for each participant. A d' score of 0 indicates that the detectability was at a chance level. The index was tested against the value 0 by a one-sample t -test. Significant results indicate a certain degree of awareness of the CS-US pairings. To further investigate participants' confidence level with their responses, we conducted an analysis on the confidence ratings according to Wicken's method (2001). To estimate the signal in each confidence category, we combined the response options into hits and false alarms with the following step: Initially all except one response options were interpreted as a "hit" (i.e., 3 to 4 were coded as a "hit"), the number of hits was reduced until the last category (4) was coded as a hit. With eight possible response options (from 4 to 4, without 0) in the present task, seven confidence categories are formed. Subsequently, d' scores were determined for each category of response, and one-sample t -tests were applied to test its significance against zero.

For the Pair Comparison Task, participants' responses to CS+ and CS- faces were summed respectively. A paired-sample t -test was employed to compute the preference rating between the CS+ and the CS-.

Imaging task

fMRI preprocessing and processing

Image preprocessing was carried out using FMRIPREP, an fMRI preprocessing pipeline recommended by Esteban et al. (2019). Each T1-weighted volume (T1w) was corrected for INU (intensity non-uniformity) and skull-stripped. Spatial normalization to the ICBM 152 nonlinear asymmetrical template was performed through nonlinear registration using brain-extracted versions of both T1 weighted volume

and template. Brain tissues of cerebrospinal fluid, white matter, and gray matter were performed on the extracted T1w. Functional data were slice time corrected and motion-corrected. This was followed by co-registration to the corresponding T1w using boundary-based registration with six degrees of freedom. Motion correcting transformations, BOLD-to-T1w transformation, and T1w-to-template warp were concatenated and applied using Lanczos interpolation. Physiological noise regressors were extracted using the component-based noise correction method (Behzadi et al., 2007). ICA-based Automatic Removal Of Motion Artefacts (ICA-AROMA) was used to remove motion artefacts (Pruim et al., 2015).

Univariate analysis

First-level analysis. Functional data was first modelled at the subject level by fitting a voxel-wise General Linear Model (GLM) to the BOLD data acquired for each run. Each run was modelled separately and included the following task regressors: the time of onset of the CS, the time of onset of the US and six motion regressors. The US regressors were orthogonalised with respect to the CS+ regressors, meaning that the CS+ regressors are statistically completely independent of the US effects. Task regressors were modelled as event-related designs and convolved with a canonical gamma hemodynamic response function, using the following formula: $\text{bold signal} = \beta_1 * \text{US} + \beta_2 * \text{CSa} + \beta_3 * \text{CSb} + \beta_4 * \text{CS-} + \text{random error}$. The main contrast was to assess the potential differences between the CS+ and the CS- during Acquisition (i.e. CS+ > CS-).

Second-level (group) analysis. We focused our fMRI analysis on the predefined cortical, subcortical and cerebellum ROIs, followed by the whole-brain analysis as an exploratory analysis. The contrast of parameter estimates (COPE) images of the CS+ > CS- was entered into a group mean model using FSL's randomize with 5000 permutations. Correction for multiple comparisons was performed using FSL's threshold-free cluster enhancement (TFCE) tool. The resulting contrast maps were thresholded at $p_{\text{FWE}} < 0.05$ with a minimum cluster size of 10 voxels. *A priori* regions of interest (ROIs) analyses on different limbic regions were conducted. We have created the limbic ROIs using the Brainnetome Atlas (Fan et al., 2016), encompassing the following regions: left dlPFC (A8v_l, A9_46d_l, A9_46v_l), right dlPFC (A8v_r, A9_46d_r, A9_46v_r), left hippocampus (cHipp_l, rHipp_l), right hippocampus (cHipp_r, rHipp_r), left amygdala (lAmyg_l, mAmyg_l), right amygdala (lAmyg_r, mAmyg_r). *A priori* regions of interest (ROIs) analyses on different lobules of the cerebellum were also conducted. ROI anatomical masks were constructed based on the spatially unbiased atlas template of the cerebellum (SUIT) Atlas (Cerebellum-SUIT.nii, (Diedrichsen, 2006), encompassing all regions. ROI analyses were performed with a threshold of $k > 10$ and $p_{\text{FWE}} < 0.05$, corrected for family-wise errors within the specified ROIs using small volume correction.

Multivariate representational similarity analyses (RSA)

The purpose of the RSA was to generate metrics to estimate the degree of neural pattern similarity across each run during acquisition. To this end, we first estimated the voxelwise activation patterns of each RSA trial with a separate first-level design. For each trial, the design matrix involved 4 regressors: the onset time of the targeted CS+, the onset times of remaining CS+, of CS- and of the US. The RSA first-level analyses used unsmoothed, non-AROMA-denoised functional data such that the fine-grained spatial pattern can be preserved according to Kriegeskorte et al. (2008). We also applied the Friston 24-parameter model (Friston, Williams, Howard, Frackowiak & Turner, 1996) to regress out motion effects. Contrasts were built to estimate t -statistic images for each RSA trial. There was a total of 4 runs in the acquisition phase. For each run, 6 CS+ and 3 CS- trials were modelled, resulting in a total of 24 CS+ and 12 CS- t -statistic maps for each participant.

Once we extracted the ROI voxel values from the participant's first-level t -statistic map, we examined the similarities in the neural pattern across runs by condition (CS+ or CS-). To estimate the similarity metric,

all single trials corresponding to each run were grouped as a set. We then calculated the Pearson Correlation Coefficients among the trials for each pair of sets, and took the median of the coefficients. To test for the significance in learning across time, we fitted the data with a growth model using a linear mixed effect model (LME) with fixed factors (*run*) and a subject-specific intercept (random factor) using the following R model formula:

$$\text{similarity} \sim \text{run} + (\text{run}|\text{subject})$$

We performed an LME separately for each ROI save for the cerebellum. To our knowledge, fine-grained functional parcellation was only available for cortical regions (e.g. Brainnetome) but not for the cerebellum. Such parcellation is essential for RSA for two reasons. The available cerebellar parcellations were enormous in size, and were largely based on structural characteristics which could involve different functional modules. Involvements of functionally irrelevant voxels would suppress the similarity metrics and hamper the reliability of the metrics.

Model comparisons were performed between the full model and a null model without any fixed factor (*run*). We used a conservative threshold to adjust for multiple comparisons within each ROI (i.e. $p = .05/3 \approx 0.0167$). Statistical analysis was performed in R studio. The linear mixed-effect model was computed using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015).

Results

Contingency awareness

The CS-US matching task revealed that the CS-US detectability was low ($d' = -0.11$, $SD = 0.37$) among our participants and not different from chance level ($t(21) = -1.38$, $p = 0.180$). The detectability for female-scream-paired and male-scream-paired-faces were also low ($d' = -0.13$ and $d' = 0.03$ respectively), and were not significantly different from zero ($t(21) = 1.71$, $p = .101$ and $t(21) = 0.60$, $p = .554$),

suggesting that most participants were not able to detect the CS-US contingency in the present study.

Further analysis of signal detectability for the confidence ratings revealed similar results to the overall d' . The d' scores were not significantly different from zero across all seven confidence intervals categories, which represents the cumulative hit rates against the cumulative false alarm rates, starting from the highest confidence rating decision for targets (see Fig. 2a; category 1: $t(21) = -0.46$, $p = .648$; category 2: $t(21) = -0.98$, $p = .338$; category 3: $t(21) = 0.40$, $p = .695$; category 4: $t(21) = -1.39$, $p = .180$; category 5: $t(21) = -1.17$, $p = .256$; category 6: $t(21) = -0.65$, $p = .524$; category 7: $t(21) = 1.96$, $p = .064$). Category 1 corresponds to response options (4) while category 7 corresponds to the response options (3 to 4). Overall, these statistically insignificant d' scores suggest that the detection of CS-US pairings was at chance level for all levels of confidence ratings. Fig. 2a illustrates the signal detectability for the confidence rating in a receiver operating characteristic curve (ROC).

Consistent with the results from the CS-US matching task, participants did not report preferences towards any groups of CS after conditioning on the Pair Comparison task (see Fig. 2b; $t(21) = 0.39$, $p = .702$).

Univariate analyses: fMRI whole brain and roi analyses (contrast CS+ > CS-)

The following section reports the findings from the whole brain and ROI analyses. We were interested in neural activations related to the conditioning of multiple CS-US pairings across the four runs.

For the whole-brain analyses, relative to the CS-, CS+ evoked a widespread activation in the cortex (Fig. 3 and Table 1). These prominent regions include the cingulate gyrus, frontal medial regions, the frontal pole, and the cerebellum (I-IV).

With regards to the ROI analyses, the CS+ relative to the CS- evoked higher activation in the bilateral dlPFC and the right cerebellum (I-IV). The results of univariate analyses in the cerebellum across each run was illustrated in supplementary Table 1.

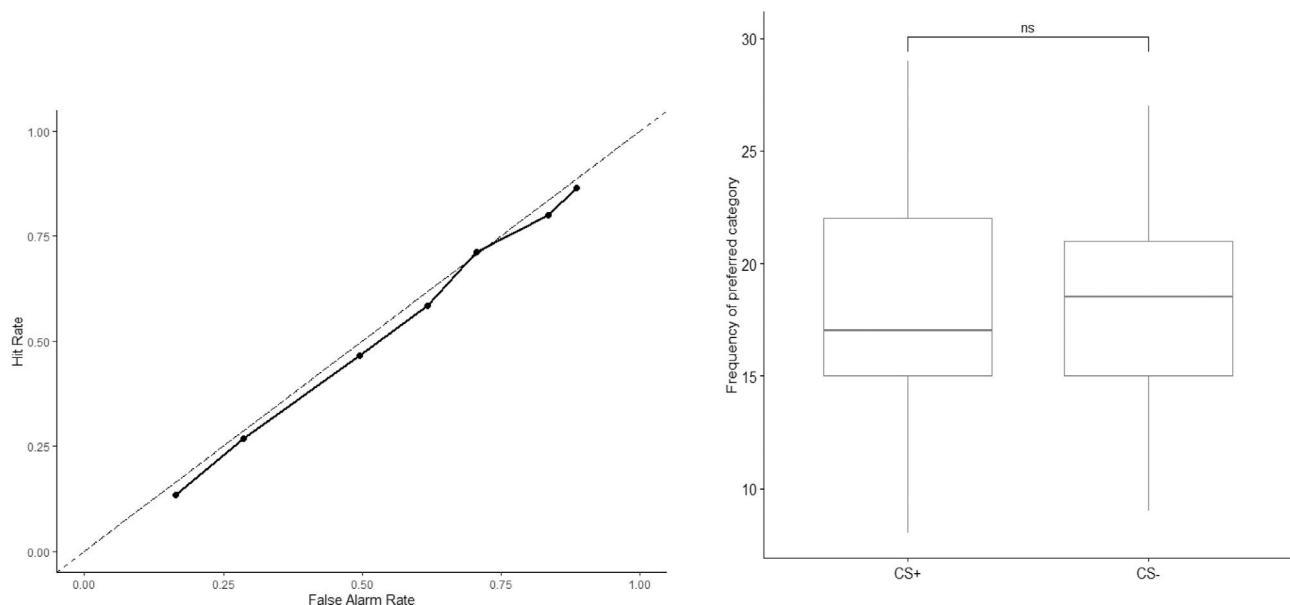
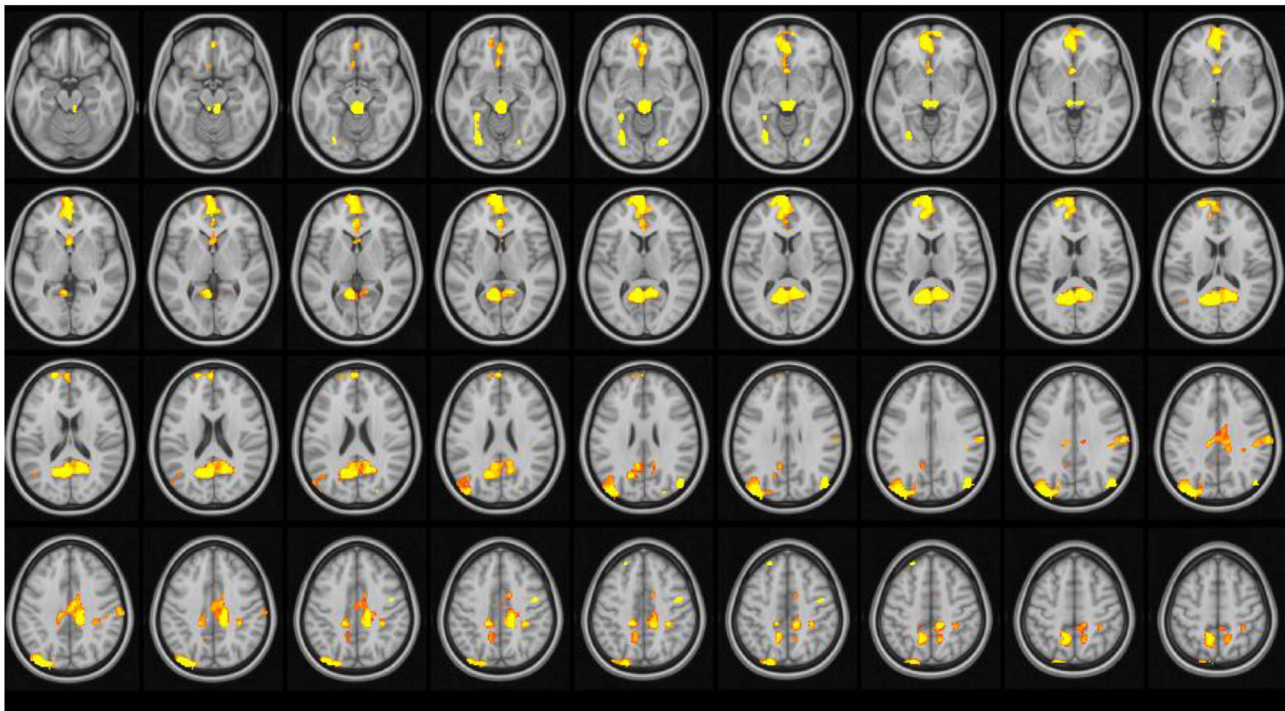


Fig. 2. a) Results of explicit CS/US matching task after Acquisition: Receiver Operating Characteristic (ROC) curve delineating the Hit and False Alarm rates for an 8-point confidence scale (−4 (surely no pairing has taken place) to 4 (surely pairing has taken place)). Successive points on the ROC are the cumulative hit rates plotted against the cumulative false alarm rate, starting with the highest confidence decision for targets. The solid line depicts scores from the present sample; the dashed line illustrates the values for a signal detectability of chance level (i.e. $d' = 0$). The d' scores were not significantly different from zero across all confidence interval categories, suggesting that the detection of CS-US pairings was at chance level for all levels of confidence ratings. b) Results of implicit Pair Comparison Task after Acquisition: On this task, a pair of CS+ and CS- faces were shown side by side and participants were asked to indicate their preference to either face. A higher score on y-axis suggests a preference towards a particular type of CS (CS+ or CS-).

a)



b)

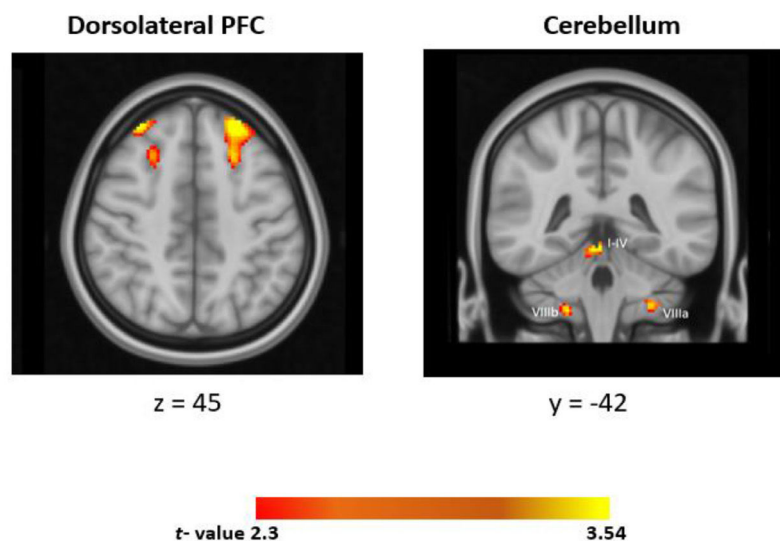


Fig. 3. Univariate differential activations during acquisition in the threat circuitry a) whole brain analysis, b) ROI analysis (CS+ > CS-). All statistical maps represent group results after mixed effect analysis ($z > 2.3$, cluster-corrected $p < 0.05$).

Multivariate representational similarity analyses

Based on the univariate result, we applied RSA to show how conditioned fear representations of CS+ and CS- changed during the course of acquisition in the bilateral dlPFC. A significant effect of *run* was found in the left dlPFC ($A9_46v_l$, $t = 13.18$, $p < .001$), the right dlPFC ($A9_46v_r$, $t = 9.51$, $p < .001$) for CS+, suggesting that CS+ exhibited an increase in neural pattern similarity to the final run as conditioning took place. CS- did not exhibit such neural similarity pattern in the left dlPFC ($t = 1.10$, $p = .279$) and the right dlPFC ($t = 0.001$, $p = .989$) (Fig. 4).

Despite the univariate analysis did not provide evidence for the involvement of the amygdala and hippocampus during conditioning, we observed an increase in neural pattern similarity in the bilateral

hippocampus, corresponding to the Brainnetome regions left rostral hippocampus (CS+: $t = 3.58$, $p = .0019$; CS-: $t = 0.08$, $p = .933$), the right rostral hippocampus (CS+: $t = 2.93$, $p = .007$; CS-: $t = 1.91$, $p = .069$). The results of the growth models of each ROI are reported in Table 2.

Discussion

The current study investigated the neural correlates of implicit associative learning using the multi-CS conditioning paradigm. Behaviourally, our results suggested that participants were not aware of the CS-US contingency after acquisition. On the neural level, our univariate results showed support for the recruitment of canonical threat-related regions including the cingulate cortex, the dorsolateral prefrontal cortex

Table 1

Acquisition: peak coordinates for the univariate activation and connectivity results across four runs (CS+ > CS-).

Analysis	structure	side	size (voxels)	x	y	z	Zmax	p FWE
Whole brain	Cingulate gyrus (posterior)	L	2472	-10	-22	34	3.54	0.0296
	Frontal medial cortex; cingulate gyrus (anterior)	R	2208	2	46	-16	3.54	0.0304
	Cerebellum I-IV	L	1954	-6	-52	0	3.54	0.0034
	Lateral Occipital cortex (superior division)	L	1074	-40	-88	24	3.54	0.0192
	Cerebellum I-IV	R	331	8	-38	-18	3.54	0.0362
	Postcentral gyrus	R	221	34	-34	38	3.54	0.0322
	Parietal operculum cortex	R	208	46	-30	30	3.54	0.0422
	Occipital fusiform gyrus	L	183	-26	-78	-14	3.54	0.0246
	Lateral occipital cortex (superior division)	R	127	48	-80	26	3.54	0.0482
	Precentral gyrus	R	47	44	-8	40	3.54	0.048
	Occipital fusiform gyrus	R	35	26	-78	-12	3.54	0.0454
	Postcentral gyrus	R	34	38	-36	72	3.54	0.0444
	Frontal pole	L	21	-22	40	44	3.54	0.0478
	Superior parietal lobule	L	10	-38	-46	62	3.54	0.0448
ROI	Dorsolateral PFC	R	344	34	40	46	3.54	<0.0001
				24	34	28	3.35	
				26	22	48	3.35	
				22	26	40	3.04	
	Dorsolateral PFC	L	712	30	42	38	2.99	<0.001
				-22	22	36	3.54	
				-26	-36	40	3.54	
				-20	40	42	3.54	
	Cerebellum I-IV	R	46	-24	32	48	3.54	0.0345
				8	-40	16	3.54	
				2	-40	16	3.54	

(dlPFC), as well as the cerebellum during acquisition. Further analyses using multivariate representational similarity analyses revealed learning-dependent changes in the bilateral dlPFC. These findings support the hypothesis that both the dlPFC and the cerebellum are implicated in largely implicit associative threat learning processes, that are characterized by very limited or even absent contingency awareness.

The dlPFC and implicit threat learning

The observed neural activation during acquisition is consistent with previous research findings that bilateral dorsolateral prefrontal cortex is implicated in fear learning (Fullana et al., 2016; LaBar & Cabeza, 2006). Importantly, the increased similarity of neural activity patterns in the bilateral dlPFC observed in our study may not only indicate learning-dependent changes in this brain region but also a cortical involvement in the early processing of threat (Pessoa & Adolphs, 2010). Although some higher-order theories propose that only conscious fear perception involves higher-order regions such as the dlPFC and the ventromedial PFC (Lau & Rosenthal, 2011; Odegaard, Knight & Lau, 2017), our findings concur with previous multi-CS conditioning studies using high temporally resolved magnetoencephalography (Rehbein et al., 2014; Roesmann et al., 2019) and a meta-analysis on unaware facial subliminal presentations (Brooks et al., 2012). This dlPFC activity appears to exercise a top-down influence modulating the perceptual processing in the ventral visual stream (Keuper, Terrighena, Chan, Junghoefer, & Lee, 2018; Roesmann et al., 2019).

Notably, the multivariate analyses provided additional information beyond univariate analyses in terms of *when* the learning took place. In addition to the differential activation of bilateral dlPFC during acquisition observed in the univariate analyses, the multivariate similarity pattern analyses unveiled the subtle but increasing neural representation over the course of learning. Our findings support that the dlPFC was recruited early in the acquisition of threat association and was increasingly engaged throughout the course of learning in the multi-CS conditioning paradigm. While the dlPFC does not directly project to the amygdala, it may control amygdala activity indirectly through its projections to the vmPFC and the lateral temporal cortex (Buhle et al., 2014; Ochsner & Gross, 2005). Moreover, the dlPFC is associated with

working memory (Curtis & D'Esposito, 2003) and selective attention (Gladwin, denUyl, Fregni & Wiers, 2012; Wilkins, Shallice & McCarthy, 1987), as well as threat appraisal (Staudinger, Erk & Walter, 2011), processes which are involved in threat learning.

Contrary to our hypothesis, the univariate analyses (both whole-brain and ROIs) did not provide evidence for significant involvement of the amygdala and hippocampus when comparing the CS+ and CS- contrast. Two possible reasons may account for this lack of differential subcortical activation in the present study. First, there is increasing evidence that CS- is not entirely neutral; the non-occurrence of the US following the CS- might position CS- as a learned safety cue involving associative learning processes (Lonsdorf et al., 2017). Given amygdala and hippocampus are key regions in safety learning (Grasser & Jovanovic, 2021), we might thus not observe the differential conditioning. Accordingly, a previous meta-analysis consisting of 27 studies on human fear conditioning did not identify robust involvement of the amygdala region (Fullana et al., 2016). Nevertheless, we observed some increasing signal similarity in the rostral regions of the hippocampus and this signal similarity was only identified in the CS+. The multivariate RSA revealed the subtle learning process that might be lost during the univariate analyses where we compared the contrast of the averaged neural activation of CS+ to that of CS- across the four runs. RSA provided further information on how one brain region responds to one category (in our case, the CS+) more strongly and consistently across time. In this regard, future studies may also consider incorporating RSA techniques to explore the learning process in fear conditioning (Visser et al., 2016).

The cerebellum and implicit threat learning

Our results are consistent with some previous findings, indicating the involvement of the cerebellum in human threat learning. Specifically, cerebellar activation of Lobules IV, VIIIa and VIIIb during the presentation of conditioned stimuli is largely in accordance with the fear conditioning study reported by Ernst and co-workers (Ernst et al., 2019) and the meta-analytic study of the involvement of these regions in emotional processing and working memory tasks (Keren-Happuch, Chen, Ho & Desmond, 2014). It was further proposed that lobules VIII, along with VIIb and IX, were associated with processing negative emotions such as fear

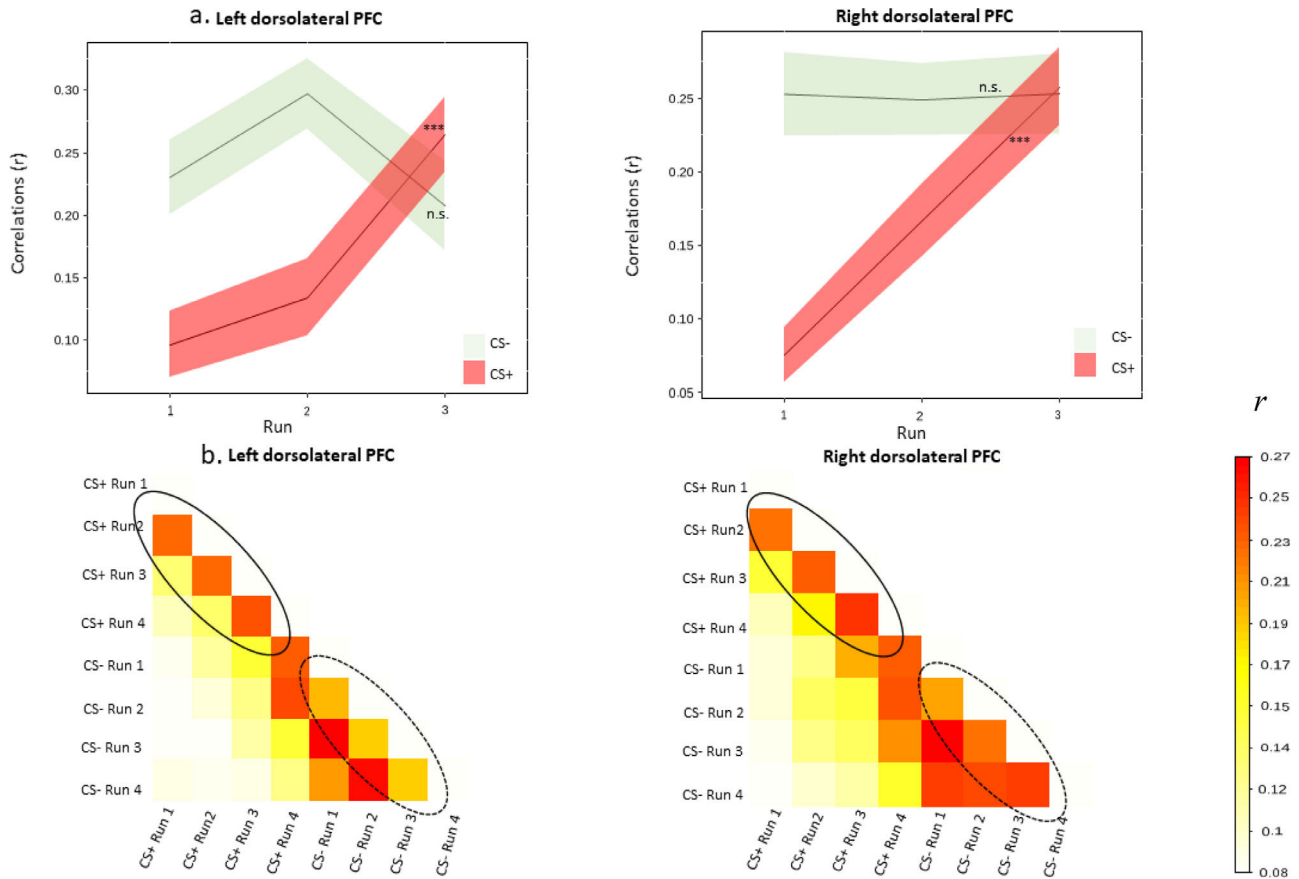


Fig. 4. (a) Similarity metrics in the dorsolateral prefrontal cortex (corresponding Brainetome locations: A9_46v_l and A9_46v_r) during acquisition. The similarity metrics in each run were compared to the final run of acquisition (i.e. run 4). As learning progressed, greater increase for neural similarity was found for the CS +, but not for the CS-. Shaded area represents SEM. *** $p < 0.001$; n.s., not significant. (b) Half of the 8×8 correlation matrix was created for the dorsolateral prefrontal cortex. The off-diagonal represents the correlations between the corresponding type of CS (CS+ or CS-) and the runs. The solid-line ellipse shows increasing correlations between runs of the CS+, while the dash-line ellipse shows relatively stable correlations between runs of the CS-. The color bar denotes the correlation coefficients across different runs and conditions.

Table 2

Growth model results for each region within the ROIs.

ROI	Left or Right	Brainetome atlas	Conditions	t-value	p-value (uncorrected)
dlPFC	Left	A8v_l_l	CS +	7.05	<0.0001
			CS-	4.82	0.000
		A9_46d_l	CS +	8.01	<0.0001
			CS-	5.52	<0.0001
		A9_46v_l	CS +	13.18	<0.0001
			CS-	-1.10	0.279
	Right	A8v_l_r	CS +	11.46	<0.0001
			CS-	0.27	0.789
		A9_46d_r	CS +	9.51	<0.0001
			CS-	0.01	0.989
Amygdala	Left	lAmyg_l	CS +	14.19	<0.0001
			CS-	0.73	0.472
		mAmyg_l	CS +	7.40	<0.0001
			CS-	2.68	0.012
	Right	lAmyg_r	CS +	1.82	0.084
			CS-	1.05	0.300
		mAmyg_r	CS +	0.10	0.918
			CS-	4.99	0.000
Hippocampus	Left	rHipp_l	CS +	2.32	0.031
			CS-	3.37	0.002
			CS-	3.58	0.002
	Right	rHipp_r	CS +	0.08	0.933
			CS-	2.93	0.007
			CS-	1.91	0.069

dlPFC: dorsolateral prefrontal cortex.

recognition (Thomasson et al., 2019). It is important to note that only a few human fear conditioning studies to-date have investigated the cerebellar region in detail, and that threat conditioning is a complex process. Neural response to aversive stimuli involves likely both motor and non-motor cerebellar circuits that control autonomic, sensorimotor, cognitive and emotions. For instance, when facing aversive stimuli, participants may want to withdraw or prepare to move, thus the anterior lobe of the cerebellum encompassing lobules I to V and lobules VI and secondary sensorimotor area in lobule VIII are likely to be implicated. Because of the large number of pairings in the multi-CS conditioning paradigm, higher cognitive functions including working memory are likely to be engaged, which are subserved by lobule VI, Crus I and VIIB (Schmahmann, 2019). It has been proposed that working memory is an essential ingredient in contingency awareness, and unaware and aware participants might show different cerebellar activations (Ernst et al., 2019). A recent investigation of threat acquisition using a 7T scanner, however, did not reveal significant cerebellar activations comparing CS+ and CS- (Batsikadze et al., 2022). Future multi-CS studies may test the robustness of cerebellar activations as well as its relation to contingency awareness.

Whole-brain analyses using the multi-CS conditioning paradigm

We ran an exploratory whole-brain analysis in addition to a ROI-based analysis to explore the neuronal activities associated with implicit threat learning. Our results demonstrate a widespread activation of the

cingulate cortex, parts of the cerebellum, superior occipital cortex and occipital fusiform gyrus, postcentral gyrus, and parietal operculum cortex. Similar to the explicit, classical conditioning paradigm, the multi-CS conditioning paradigm recruits the cerebellum, the anterior and middle cingulate cortex, and the parietal operculum/somatosensory cortex. These regions are known to show significant activations during fear acquisition and expression (Fullana et al., 2020, 2018; Milad & Quirk, 2012). It is perhaps interesting to note that the occipital region, including the fusiform gyrus, was activated to a greater extent by the CS+ compared to the CS- in the multi-CS conditioning paradigm. This finding concurs with several MEG multi-CS studies of neutral faces, which observed early learning effects in the occipital areas after stimulus onset (Rehbein et al., 2014; Steinberg et al., 2013)

Threat learning and contingency awareness

Whether consciousness is needed to acquire threat conditioning is a heated debate (Lovibond & Shanks, 2002; Mertens & Engelhard, 2020; Mitchell, DeHouwer & Lovibond, 2009), but it is generally agreed that conscious experience of fear is not entirely independent from its unconscious physiological processes. Threat learning can be considered as a product of cortical circuits that underlie working memory and other cognitive functions, as well as subcortical circuits that control physiological responses and defensive behaviours (LeDoux & Pine, 2016). While the subcortical circuits operate without awareness, the cortical circuits receive inputs from the subcortical circuits, integrate them, and form the conscious feeling of fear. The findings of the current study are likely to support the dissociation between conscious and unconscious processes of fear learning: the dlPFC and cerebellar activations during conditioning were evident and this was independent of participants' explicit awareness of the CS-US associations. We believe that acknowledging both processes in threat learning could advance our understanding of the mechanisms underlying anxiety or fear-related disorders, guiding us to develop innovative interventions that target maladaptive fear and anxiety at the implicit level.

Limitations

First, the sample size of the study was small. Findings from the univariate and multivariate analyses with a small sample size are prone to type I and type II errors (Marek et al., 2022; Turner, Paul, Miller, & Barbey, 2018). Furthermore, univariate and multivariate analyses have distinct assumptions and methods in controlling for the false positive errors. Multivariate analyses are generally more sensitive than univariate ones, producing more stable and reliable results (Grady, Rieck, Nichol, Rodrigue & Kennedy, 2021). Given the greater sensitivity of multivariate analyses, fewer participants might be required. However, it is also important to note that other factors such as within- vs between-subject design, duration and nature of the task, individual-level data will also affect the replicability of fMRI results (Nee, 2019). Future studies should aim to replicate our findings with a larger sample sizes in order to corroborate and extend findings of this report. Second, the present report employed a 100% reinforcement schedule in the acquisition phase, i.e. all CS+ were, though with variable SOAs, followed by the US during the conditioning phase. Responses to the CS+ could thus include US processing as a potential confound. To reduce the impact of this confound, we included a regressor for the US events to minimize their influence on the beta estimates for the regressors of interest in the GLM. We made the CS+ and CS- orthogonalised to the US in our models, hence the CS effects are completely and statistically independent of the US. Third, we employed offline measures to infer the awareness of the CS-US associations. It was argued that offline measures of contingency awareness reflect the memory or recognition of the learning per se. Offline rating was preferred in the present study as continuous online rating might boost the explicit learning process, interfering with the process of implicit learning (Lonsdorf et al., 2017).

Conclusion

Our findings demonstrated the implication of the dlPFC and the cerebellum in threat learning using the multi-CS conditioning paradigm. Importantly, threat learning was acquired under the conditions where the CS-US contingency awareness was highly limited or even absent. Our study contributes to the current debate on the role of contingency awareness during fear learning and extends the investigation of implicit learning to the canonical regions of the fear circuitry: the dorsolateral prefrontal cortex and the cerebellum.

Declaration of Competing Interest

none

Funding

The study is supported by the Germany/Hong Kong Joint Research Scheme (G-HKU707/17) and DAAD: German Academic Exchange Service (Project: 57390152) to K. Roesmann and The Early Career Scheme (27607921) to C.L.M. Lam.

Acknowledgement

We thank Professor Tatia M.C. Lee for helpful discussions and comments on the experiment and the earlier draft of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijchp.2022.100357.

References

- Andreatta, M., Glotzbach-Schoon, E., Mühlberger, A., Schulz, S. M., Wiemer, J., & Pauli, P. (2015). Initial and sustained brain responses to contextual conditioned anxiety in humans. *Cortex; a journal devoted to the study of the nervous system and behavior*, 63, 352–363. doi:10.1016/j.cortex.2014.09.014.
- Bach, D. R., Weiskopf, N., & Dolan, R. J. (2011). A stable sparse fear memory trace in human amygdala. *Journal of Neuroscience*, 31(25), 9383–9389. doi:10.1523/JNEUROSCI.1524-11.2011.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.
- Batsikadze, G., Diekmann, N., Ernst, T. M., Klein, M., Maderwald, S., & Deuschl, C. (2022). The cerebellum contributes to context-effects during fear extinction learning: A 7T fMRI study. *NeuroImage*, 253, 119080. doi:10.1016/j.neuroimage.2022.119080.
- Brockelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghofer, M. (2011). Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing: Magnetoencephalographic Correlates. *Journal of Neuroscience*, 31(21), 7801–7810. doi:10.1523/JNEUROSCI.6236-10.2011.
- Brooks, S. J., Savov, V., Allzén, E., Benedict, C., Fredriksson, R., & Schiöth, H. B. (2012). Exposure to subliminal arousing stimuli induces robust activation in the amygdala, hippocampus, anterior cingulate, insular cortex and primary visual cortex: A systematic meta-analysis of fMRI studies. *NeuroImage*, 59(3), 2962–2973. doi:10.1016/j.neuroimage.2011.09.077.
- Buhle, J. T., Silvers, J. A., Wage, T. D., Lopez, R., Onyemkwo, C., Kober, H., et al. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex*, 24(11), 2981–2990. doi:10.1093/cercor/bht154.
- Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423. doi:10.1016/S1364-6613(03)00197-9.
- Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. *NeuroImage*, 33(1), 127–138. doi:10.1016/j.neuroimage.2006.05.056.
- Ernst, T.M., Brol, A., Gratz, M., Ritter, C., Bingel, U., Schlamann, M. et al. (2019). The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. *BioRxiv*, 1–26. https://doi.org/10.1101/600494
- Fanselow, M. S. (2000). Contextual fear, gestalt memories, and the hippocampus. *Behavioural Brain Research*, 110(1–2), 73–81. doi:10.1016/S0166-4328(99)00186-2.
- Feinstein, J., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, 21(1), 34–38. doi:10.1016/j.cub.2010.11.042.The.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35(3), 346–355. doi:10.1002/mrm.1910350312.

- Fullana, M. A., Dunsmoor, J. E., Schruers, K. R. J., Savage, H. S., Bach, D. R., & Harrison, B. J. (2020). Human fear conditioning: From neuroscience to the clinic. *Behaviour Research and Therapy*, 124, 103528. doi:10.1016/j.brat.2019.103528.
- Fullana, Miquel A., Albajes-Eizaguirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., et al. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews*, 88, 16–25. doi:10.1016/j.neubiorev.2018.03.002.
- Fullana, Miquel A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Ávila-Parcet, A., et al. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, 21(4), 500–508. doi:10.1038/mp.2015.88.
- Gladwin, T. E., denUyl, T. E., Fregni, F. F., & Wiers, R. W. (2012). Enhancement of selective attention by tDCS: Interaction with interference in a Sternberg task. *Neuroscience Letters*, 512(1), 33–37. doi:10.1016/j.neulet.2012.01.056.
- Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M., & Kennedy, K. M. (2021). Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. *Human Brain Mapping*, 42(1), 204–219. doi:10.1002/hbm.25217.
- Grasser, L. R., & Jovanovic, T. (2021). Safety learning during development: Implications for development of psychopathology. *Behavioural Brain Research*, 408, 113297. doi:10.1016/j.bbr.2021.113297.
- Greco, J. A., & Liberzon, I. (2016). Neuroimaging of Fear-Associated Learning. *Neuropsychopharmacology*, 42(2), 782–790. doi:10.1038/npp.2015.255.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. U.S.: Wiley.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. doi:10.1016/j.neuroimage.2011.09.015.
- Kalisch, R., & Gerlicher, A. M. V. (2014). Making a mountain out of a molehill: On the role of the rostral dorsal anterior cingulate and dorsomedial prefrontal cortex in conscious threat appraisal, catastrophizing, and worrying. *Neuroscience and Biobehavioral Reviews*, 42, 1–8. doi:10.1016/j.neubiorev.2014.02.002.
- Kattoor, J., Thürling, M., Gizewski, E. R., Forsting, M., Timmann, D., & Elsenbruch, S. (2014). Cerebellar contributions to different phases of visceral aversive extinction learning. *Cerebellum (London, England)*, 13(1), 1–8. doi:10.1007/s12311-013-0512-9.
- Keren-Happuch, E., Chen, S.-H. A., Ho, M.-H. R., & Desmond, J. E. (2014). A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Human Brain Mapping*, 35(2), 593–615. doi:10.1002/hbm.22194.
- Keuper, K., Terrighena, E. L., Chan, C. C. H., Junghoefer, M., & Lee, T. M. C. (2018). How the Dorsolateral Prefrontal Cortex Controls Affective Processing in Absence of Visual Awareness – Insights From a Combined EEG-rTMS Study. *Frontiers in Human Neuroscience*, 12, 412. doi:10.3389/fnhum.2018.00412.
- King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B., & Diedrichsen, J. (2019). Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature Neuroscience*, 22(8), 1371–1378. doi:10.1038/s41593-019-0436-x.
- Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4. doi:10.3389/neuro.06.004.2008.
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1), 54–64. doi:10.1038/nrn1825.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373. doi:10.1016/j.tics.2011.05.009.
- Ledoux, J., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, 19(5), 269–282. doi:10.1038/nrn.2018.22.
- LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *American Journal of Psychiatry*, 173(11), 1083–1093. doi:10.1176/appi.ajp.2016.16030353.
- Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., et al. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews*, 77, 247–285. doi:10.1016/j.neubiorev.2017.02.026.
- Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28(1), 3–26. doi:10.1037/0097-7403.28.1.3.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9*. ISBN 91-630-7164-9.
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), 654–660. doi:10.1038/s41586-022-04492-9.
- Maren, S., & Holmes, A. (2016). Stress and fear extinction. *Neuropsychopharmacology*, 41, 58–79. doi:10.1038/npp.2015.180.
- Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, 14(6), 417–428. doi:10.1038/nrn3492.
- Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neuroscience & Biobehavioral Reviews*, 108, 254–268. doi:10.1016/j.neubiorev.2019.11.012.
- Milad, M. R., & Quirk, G. J. (2012). Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annual Review of Psychology*, 63(1), 129–151. doi:10.1146/annurev.psych.121208.131631.
- Mitchell, C. J., DeHouwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2), 183–246. doi:10.1017/S0140525X09000855.
- Nee, D. E. (2019). fMRI replicability depends upon sufficient individual-level data. *Communications Biology*, 2(1), 1–4. doi:10.1038/s42003-019-0378-6.
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–249. doi:10.1016/j.tics.2005.03.010.
- Odegaard, B., Knight, R. T., & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *Journal of Neuroscience*, 37(40), 9593–9602. doi:10.1523/JNEUROSCI.3217-16.2017.
- Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a “low road” to “many roads” of evaluating biological significance. *Nature Reviews Neuroscience*, 11(11), 773–783. doi:10.1038/nrn2920.
- Pierce, J. E., & Péron, J. (2020). The basal ganglia and the cerebellum in human emotion. *Social Cognitive and Affective Neuroscience*, 15(5), 599–613. doi:10.1093/scan/nsaa076.
- Rehbein, M. A., Steinberg, C., Wessing, I., Pastor, M. C., Zwitterlood, P., & Keuper, K. (2014). Rapid Plasticity in the Prefrontal Cortex during Affective Associative Learning. *PLoS one*, 9(10), e110720. doi:10.1371/journal.pone.0110720.
- Roesmann, K., Dellert, T., Junghoefer, M., Kissler, J., Zwitterlood, P., Zwanzger, P., et al. (2019). The causal role of prefrontal hemispheric asymmetry in valence processing of words – Insights from a combined cTBS-MEG study. *NeuroImage*, 191, 367–379. doi:10.1016/j.neuroimage.2019.01.057.
- Roesmann, K., Wiens, N., Winker, C., Rehbein, M. A., Wessing, I., & Junghoefer, M. (2020). Fear generalization of implicit conditioned facial features – Behavioral and magnetoencephalographic correlates. *NeuroImage*, 205, 116302. doi:10.1016/j.neuroimage.2019.116302.
- Schmahmann, J. D. (2019). The cerebellum and cognition. *Neuroscience Letters*, 688, 62–75. doi:10.1016/j.neulet.2018.07.005.
- Sparta, D. R., Smithuis, J., Stamatakis, A. M., Jennings, J. H., Kantak, P. A., Ung, R. L., et al. (2014). Inhibition of projections from the basolateral amygdala to the entorhinal cortex disrupts the acquisition of contextual fear. *Frontiers in Behavioral Neuroscience*, 8, 6–11. doi:10.3389/fnbeh.2014.00129.
- Staudinger, M. R., Erk, S., & Walter, H. (2011). Dorsolateral Prefrontal Cortex Modulates Striatal Reward Encoding during Reappraisal of Reward Anticipation. *Cerebral Cortex*, 21(11), 2578–2588. doi:10.1093/cercor/bhr041.
- Steinberg, C., Bröckelmann, A.-K., Rehbein, M., Döbel, C., & Junghöfer, M. (2013). Rapid and highly resolving associative affective learning: Convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological Psychology*, 92(3), 526–540. doi:10.1016/j.biopsycho.2012.02.009.
- Steinberg, C., Döbel, C., Schupp, H. T., Kissler, J., Elling, L., Pantev, C., et al. (2012). Rapid and highly resolving: Affective evaluation of olfactorily conditioned faces. *Journal of Cognitive Neuroscience*, 24(1), 17–27. doi:10.1162/jocn_a.00067.
- Stoodley, C. J., & Schmahmann, J. D. (2018). Functional topography of the human cerebellum. (1st ed.). *Handbook of clinical neurology: 154* New York, NY: Elsevier B.V.
- Strata, P. (2015). The emotional cerebellum. *Cerebellum (London, England)*, 14(5), 570–577. doi:10.1007/s12311-015-0649-9.
- Thomasson, M., Saj, A., Benis, D., Grandjean, D., Assal, F., & Péron, J. (2019). Cerebellar contribution to vocal emotion decoding: Insights from stroke and neuroimaging. *Neuropsychologia*, 132, (July) 107141. doi:10.1016/j.neuropsychologia.2019.107141.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3), 242–249. doi:10.1016/j.psychres.2008.05.006.
- Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, 1, 62. doi:10.1038/s42003-018-0073-z.
- Visser, R. M., deHaan, M. I. C., Beemsterboer, T., Haver, P., Kindt, M., & Scholte, H. S. (2016). Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology*, 53(8), 1117–1127. doi:10.1111/psyp.12665.
- Wickens, T. D. (2001). *Elementary signal detection theory*. New York: Oxford University Press.
- Wilkins, A. J., Shallice, T., & McCarthy, R. (1987). Frontal lesions and sustained attention. *Neuropsychologia*, 25(2), 359–365. doi:10.1016/0028-3932(87)90024-8.