Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

The cerebellum contributes to context-effects during fear extinction learning: A 7T fMRI study

Giorgi Batsikadze[a,1,*], Nicolas Diekmann[b,1], Thomas Michael Ernst[a,c], Michael Klein[a], Stefan Maderwald[c], Cornelius Deuschl[d], Christian Josef Merz[e], Sen Cheng[b], Harald H. Quick[c,f], Dagmar Timmann[a,c]

[a] *Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), Essen University Hospital, University of Duisburg-Essen, Hufelandstraße 55, Essen 45147, Germany*
[b] *Institute for Neural Computation, Ruhr University Bochum, Bochum, Germany*
[c] *Erwin L. Hahn Institute for Magnetic Resonance Imaging, University of Duisburg-Essen, Essen, Germany*
[d] *Institute of Diagnostic and Interventional Radiology and Neuroradiology, Essen University Hospital, Essen, Germany*
[e] *Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Ruhr University Bochum, Bochum, Germany*
[f] *High-Field and Hybrid MR Imaging, Essen University Hospital, Essen, Germany*

ARTICLE INFO

Keywords:
Fear conditioning
Acquisition
Associative learning
Erasure
Inhibitory learning
Human

ABSTRACT

The cerebellum is involved in the acquisition and consolidation of learned fear responses. Knowledge about its contribution to extinction learning, however, is sparse. Extinction processes likely involve erasure of memories, but there is ample evidence that at least part of the original memory remains. We asked the question whether memory persists within the cerebellum following extinction training. The renewal effect, that is the reoccurrence of the extinguished fear memory during recall in a context different from the extinction context, constitutes one of the phenomena indicating that memory of extinguished learned fear responses is not fully erased during extinction training. We performed a differential AB-A/B fear conditioning paradigm in a 7-Tesla (7T) MRI system in 31 young and healthy men. On day 1, fear acquisition training was performed in context A and extinction training in context B. On day 2, recall was tested in contexts A and B. As expected, participants learned to predict that the CS+ was followed by an aversive electric shock during fear acquisition training. Skin conductance responses (SCRs) were significantly higher to the CS+ compared to the CS- at the end of acquisition. Differences in SCRs vanished in extinction and reoccurred in the acquisition context during recall indicating renewal. Fitting SCR data, a deep neural network model was trained to predict the correct shock value for a given stimulus and context. Event-related fMRI analysis with model-derived prediction values as parametric modulations showed significant effects on activation of the posterolateral cerebellum (lobules VI and Crus I) during recall. Since the prediction values differ based on stimulus (CS+ and CS-) and context during recall, data provide support that the cerebellum is involved in context-related recall of learned fear associations. Likewise, mean β values were highest in lobules VI and Crus I bilaterally related to the CS+ in the acquisition context during early recall. A similar pattern was seen in the vermis, but only on a trend level. Thus, part of the original memory likely remains within the cerebellum following extinction training. We found cerebellar activations related to the CS+ and CS- during fear acquisition training which likely reflect associative and non-associative aspects of the task. Cerebellar activations, however, were not significantly different for CS+ and CS-. Since the CS- was never followed by an electric shock, the cerebellum may contribute to associative learning related to the CS, for example as a safety cue.

1. Introduction

The cerebellum plays an important role in motor control and motor learning (Manto et al., 2012) and is thought to be crucial in the predictive control of movements (Popa and Ebner, 2019). Neuroimaging and patient data provide good evidence that the cerebellum is equally involved in cognitive and emotional processes in humans (Diedrichsen et al., 2019; Guell and Schmahmann, 2020; Schmahmann, 2019). Likewise, there is increasing evidence that the cerebellum is involved in processing of predictions and prediction errors in the cognitive and emotional domains (Hull, 2020; Sokolov et al., 2017). One important emotion for survival is fear. The cerebellum has known anatomical and functional connections with many parts of the

[*] Corresponding author.
E-mail address: giorgi.batsikadze@uk-essen.de (G. Batsikadze).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2022.119080.
Received 8 October 2021; Received in revised form 14 February 2022; Accepted 7 March 2022
Available online 9 March 2022.

neural fear network including periaqueductal gray (PAG, Frontera et al., 2020; Koutsikou et al., 2014), amygdala (Farley et al., 2016), hypothalamus (Dietrichs and Haines, 1989; Onat and Cavdar, 2003), hippocampus (Heath and Harper, 1974; Liu et al., 2012; Newman and Reza, 1979), and prefrontal cortex (Middleton and Strick, 2001; Watson et al., 2014). Being able to learn to predict potentially harmful or threatening events is essential for survival, and can be assessed in the laboratory using fear conditioning paradigms (Graham and Milad, 2013). Acquisition and extinction of learned fear responses is thought to be driven by prediction errors (Holland and Schiffino, 2016; Rescorla and Wagner, 1972), that is the unexpected occurrence of the unconditioned stimulus in initial acquisition trials, and its unexpected omission in initial extinction trials, respectively. There is increasing evidence that the cerebellum plays a role in the processing of aversive predictions and prediction errors (Apps and Strata, 2015; Ernst et al., 2019), which may be related to timing (Frontera et al., 2020).

Animal and human studies have shown that the cerebellum contributes to the acquisition and consolidation of conditioned fear responses (Apps and Strata, 2015; Dubois et al., 2020; Kim and Jung, 2006): Lesions of the cerebellar vermis impede fear-conditioned bradycardia in rats and humans (Maschke et al., 2002; Supple and Leaton, 1990a, 1990b). Lesions of vermal lobule VIII disrupt conditioned freezing behavior in rats (Han et al., 2021; Koutsikou et al., 2014). Furthermore, rodent studies revealed that vermal lobules V and VI are involved in fear memory consolidation (Sacchetti et al., 2002, 2004). Moreover, neuroimaging studies show that the cerebellar vermis and parts of the cerebellar hemispheres are involved in the acquisition of conditioned fear responses in humans (Ernst et al., 2019; Fischer et al., 2000; Lange et al., 2015; Ploghaus et al., 1999).

Learning to predict potentially harmful events is important for survival, but it is equally important to extinguish previously learned associations when no longer needed (Craske et al., 2018). Lack of extinction of learned fear contributes to the pathophysiology of many types of anxiety disorders (Maren et al., 2013). As yet, the contribution of the cerebellum to extinction of learned fear responses has not been studied in detail. Extinction-related activation of cerebellar vermis and hemispheres have been reported in functional magnetic resonance imaging (fMRI) studies using somatic and visceral pain as aversive unconditioned stimuli (US) in humans (Kattoor et al., 2014; Utz et al., 2015). Cerebellar activations, however, were commonly weak and limited to the beginning of extinction training (Ernst et al., 2019). Weak or absent cerebellar activations have also been reported in neuroimaging studies of extinction of conditioned eyeblink responses (Ernst et al., 2017; Molchan et al., 1994; Thürling et al., 2015).

Extinction of learned fear responses is thought to involve both erasure and new inhibitory learning mechanisms (Barad, 2006; Myers and Davis, 2007). Learning theorists have proposed that extinction leads to the erasure of the original association (Rescorla and Wagner, 1972). The "erasure" hypothesis of extinction is supported by cellular recording studies in the cerebellar cortex showing that plastic changes related to the acquisition of conditioned responses to aversive US are reversed during extinction training. This has been shown during extinction training of conditioned eyeblink responses (Jirenhed et al., 2007) in ferrets, and during extinction training of conditioned fear responses in fish (Yoshida and Kondo, 2012). The inhibitory connection from the cerebellar nuclei to the inferior olive plays an important role for this bidirectional learning to occur in the cerebellar cortex (Hesslow and Ivarsson, 1996; Kim et al., 2020; Medina et al., 2002).

On the other hand, phenomena like spontaneous recovery, savings, reinstatement and renewal show that at least part of the memory obtained during acquisition training is maintained during extinction training (Bouton, 2004; Herry et al., 2010; Larrauri and Schmajuk, 2008; Walther et al., 2021) Pavlov (1927). was the first to suggest that extinction learning involves a newly learned inhibition of the original association. There is good evidence of an extinction network, which suppresses the expression of conditioned fear responses (Milad and Quirk, 2012). The new inhibitory learning is context-dependent and involves the ventromedial prefrontal cortex (vmPFC) and the hippocampus, with the vmPFC inhibiting amygdala activity during extinction training and extinction recall (Milad and Quirk, 2012; Phelps et al., 2004). It has been proposed that the cerebellum is also under this inhibitory control, possibly based on reduced salience of the conditioned stimulus (CS) mediated via less amygdala output (Farley et al., 2016; Hu et al., 2015; Inoue et al., 2020; Robleto et al., 2004). However, experimental evidence for extra-cerebellar inhibition of acquisition related plasticity within the cerebellum is sparse. Increased functional connectivity between vmPFC, dorsal anterior cingulate cortex and the cerebellum has been demonstrated during extinction recall in a cognitive association task (Kinner et al., 2016).

Thus, there is clear evidence that at least part of the initial associative fear memory is retained in extinction. One would expect that this is equally the case in the cerebellum, but, as yet, experimental evidence has been missing. The aim of the present study was to provide evidence that part of the original associative memory is preserved in the cerebellum during extinction of learned fear responses. We studied a two-day AB-A/B fear conditioning paradigm (Bouton and King, 1983) in a group of young healthy human participants using fMRI in a 7-Tesla (7T) MR system. Fear acquisition training was performed in context A and extinction training in context B. On a second day, recall was tested in both the acquisition context A and the extinction context B. The reoccurrence of conditioned responses in the acquisition context is called renewal effect (Bouton, 2004). In case part of the acquisition-related memory in the cerebellum was at least partly preserved, cerebellar activation was expected to return during recall in the acquisition, but not the extinction context.

Indeed, our results are consistent with the hypothesis that part of the original associative fear memory is preserved in the cerebellum following extinction training. In addition, and unexpected at first, we found a lack of differential cerebellar activation related to the partially reinforced conditioned stimulus (CS+) and the non-reinforced conditioned stimulus (CS-) during fear acquisition training. One possible explanation is that the cerebellum is not only involved in predicting harmful events, but also in associative processes related to the CS-, e.g., predicting the absence of a harmful event, i.e., safety. Cerebellar involvement in non-associative processes, however, may also play a role.

2. Methods

2.1. Participants

Only male participants were included in this study because menstrual cycle and oral contraceptives affect acquisition and extinction of conditioned fear (Merz et al., 2018). A total of 41 young and healthy men (mean age 25.93 ± 3.6 years) performed the experiment. Seven participants had to be excluded due to technical errors, e.g. volume orientation mix-ups or loss of adjustment volume settings, two due to high depression scores based on the DASS-21 questionnaire (Lovibond and Lovibond, 1995) and one due to constant motion and related artifacts throughout MRI acquisition. Thus, a total of 31 participants (mean age 26.22 ± 3.67 years) were included in final data analyses. Participants were recruited on university campuses via announcements posted on notice boards and online.

None of the participants presented with neurological or neuropsychiatric disorders based on medical history. None were taking centrally acting drugs. All participants were right-handed based on the Edinburgh handedness inventory (Oldfield, 1971) and instructed to refrain from alcohol intake at least 24 h prior to the experiment. The study was approved by the Ethics Committee of the University Hospital Essen and conforms to the principles laid down in the Declaration of Helsinki. Informed consent was obtained from all participants who were compensated with 75 Euros for their participation.

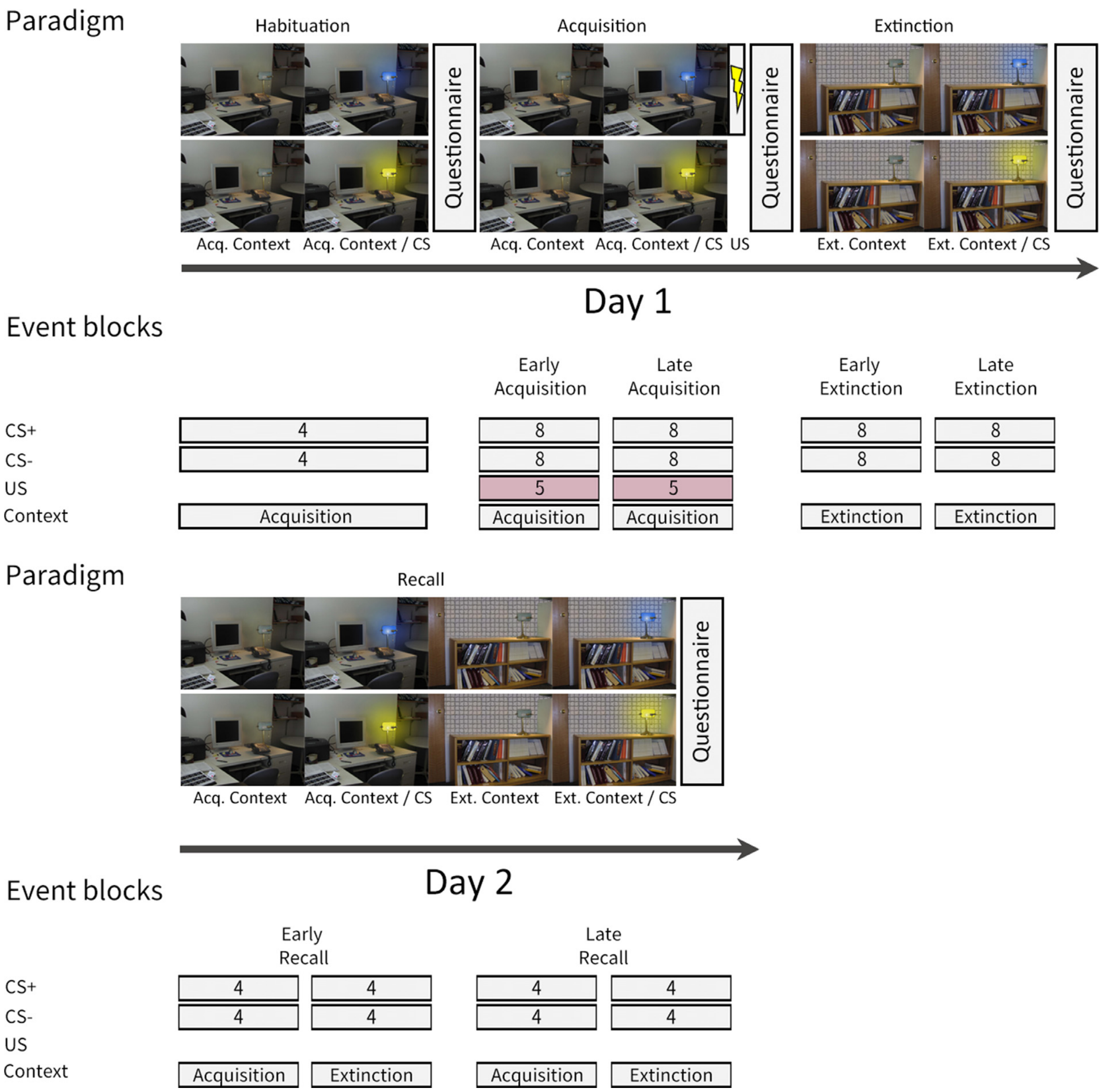


Fig. 1. Differential AB-A/B fear conditioning paradigm and fMRI event blocks (modified from Milad et al., 2007). Conditioned stimuli (CS) were presented in two contexts (acquisition context (A) and extinction context (B)): a picture of a desk ("office") or a bookshelf ("library"). CS were represented by a lamp shining either in blue or yellow color. The unconditioned stimulus (US) was an electric stimulation applied to the left shin. Fear acquisition (acq.) and extinction training (ext.) was performed on day 1, recall was tested on day 2.

2.2. *Fear conditioning paradigm*

A differential AB-A/B fear conditioning paradigm was used based on the paradigm published in Milad et al. (2007) (**Fig. 1**). CS were presented in two contexts (A, B): a picture of a desk (“office”) or a bookshelf (“library”) each containing a lamp (Milad et al., 2007). The two CS were represented by the same lamp shining either in blue or yellow color. The US was an electric shock applied to the left shin. The CS+ was followed by the US (paired CS+/US trial) during fear acquisition training in 62.5% of the cases. The CS- was never followed by the US. Participants were informed that should they perceive a pattern between CS and US presentations, the pattern would not change during the experiment.

On day 1, a brief habituation phase (4 CS+ only trials, 4 CS- only trials) presented in the acquisition context (A) was followed by fear acquisition training (10 paired CS+/US trials, 6 CS+ only trials, 16 CS-) in the acquisition context (A), and extinction training (16 CS+ only trials, 16 CS- only trials) presented in the extinction context (B). On day 2, recall was tested in the acquisition and extinction context (each containing 8 CS+ only trials and 8 CS- only trials).

Presentation of the paradigm was controlled by a computer running the software Presentation (version 20.0, Neurobehavioral System Inc., Berkeley, CA). For synchronization of stimulus presentation with fMRI acquisition, MRI scan triggers were supplied to the control computer.

Images were projected onto a rear projection screen inside the scanner bore using a standard projector. Images were visible to the participants through a mirror mounted on the radiofrequency head coil (1chTx/32ChRx coil, Nova Medical Inc. Wilmington, USA).

The electric stimulation was generated by a constant current stimulator (DS7A, Digitimer Ltd., London, UK) and applied to the left shin via a concentric (ring-shaped) bipolar surface electrode with 6 mm conductive diameter and a central platinum pin (WASP electrode, Specialty

Developments, Bexley, UK). Electrode position was marked with a permanent marker on day 1 to use the same electrode position on day 2. The 100 ms US consisted of a short train of four consecutive 500 µs current pulses (maximum output voltage: 400 V) with an inter pulse interval of 33 ms. Stimulation intensity was determined immediately before start of MRI measurements: the stimulation current was gradually increased, and participants were asked to report on the perceived sensation intensity until an "unpleasant but not painful" intensity was reached (8.19 ± 4.69 mA, range 3.36 mA - 23.04 mA). To counteract habituation to the US leading to weakening of the conditioned responses (Inoue et al., 2020), 20% was added to the individual thresholds (mean added current 2.05 mA ± 1.71 mA). The final individual current setting was kept the same for all stimulations.

Each trial consisted of a 12 s context and an 8 s CS presentation. CS started 1–3 s after context onset. Time of CS onset varied for a better distinction of CS from context. In case of reinforced trials, the US was presented 7.9 s after CS onset and co-terminated with the CS. Contexts and CS colors were pseudo-randomly counterbalanced across participants. A neutral gray background with a black cross ("fixation cross") was displayed between visual stimulus presentations (intertrial interval, ITI, randomized between 22.36 and 27.32 s). To avoid dazzle effects the ITI image was chosen to be of approximately the same luminosity as the mean luminosity of the context images.

The different trial types in each phase were presented in pseudo-randomized order with two restrictions: Firstly, the first two trials and the very last trial of fear acquisition training were always paired CS+/US trials, no more than two consecutive trials events of the same kind were shown, and the number of events of each kind was kept identical in the first and second half of each learning phase. Additionally, during recall, no more than two consecutive trials were shown in the same context, and context presentations in the first and second half of recall were counterbalanced. During fear acquisition and extinction training, the order of events was the same for all participants. Order of CS+ and CS- events during habituation was counterbalanced. During recall, the order of events was the same for all participants except for the first and the third trials: In one half of the participants, CS+ was presented in acquisition context (A) in trial 1, followed by CS+ in extinction context (B) in trial 2, and the CS- in acquisition context (A) in trial 3. In the other half of the participants, the CS- was presented in the acquisition context (A) in trial 1, followed by the CS+ in the extinction context (B) in trial 2, and the CS+ in the acquisition context (A) in trial 3.

Each experimental phase was performed within a separate session of fMRI data acquisition.

2.3. Physiological data acquisition

Throughout the experiment, skin conductance responses (SCRs), pulse and breathing rate were acquired using MRI-compatible skin conductance, pulse oximetry and differential air pressure modules, and appropriate hardware filters sampling at 1 kHz (MP160, BIOPAC Systems Inc., Goleta, CA). Two skin conductance electrodes were attached to the participant's left hypothenar, approximately 20 mm apart. The pulse oximetry sensor was clipped to the participant's left index finger. A respiratory bellows was attached to the participant's lower abdomen using a hook-and-loop belt.

2.4. Skin conductance response (SCR) evaluation

SCR data processing was performed using MATLAB software (Release 2019a, RRID:SCR_001622, The MathWorks Inc., Natick, MA). To eliminate high-frequency noise and low-frequency drifts SCR data was bandpass filtered (-0.5 to 10 Hz). Semi-automated peak detection was performed, and SCRs were defined as the maximum trough-to-peak-amplitude within a time interval from 1 to 8.5 s after CS onset (Pineles et al., 2009).

Raw SCRs were normalized through a logarithmic (LN(1+SCR)) transformation (Boucsein, 2012; Venables and Christie, 1980). Shapiro-Wilk-test was used to test the normalized data and the distribution of residuals for normality. Since the normality test revealed a non-normal distribution of SCRs and the residuals ($p < 0$ 05), data were analyzed with non-parametric statistical analysis for repeated measures using rank-based F-tests (ANOVAF option in the PROC MIXED method in SAS, SAS Studio 3.8, SAS Institute Inc, Cary, NC, USA), which has been recommended for dealing with skewed distributions, outliers or small sample sizes (Brunner et al., 1997, 2002, 1999; Shah and Madden, 2004).

Non-parametric ANOVA-type statistics for repeated measures were used separately for each phase with SCRs as dependent variable and stimulus (CS+, CS-), trial (1 to 16) and context (acquisition context or extinction context, for recall only) as within-subjects factors as well as their interactions. Throughout the manuscript, in case of significant results of non-parametric ANOVA, post-hoc comparisons were performed using least square means tests and were adjusted for multiple comparisons using the Tukey-Kramer method.

2.5. Questionnaires

Participants were required to answer four questionnaires following each phase of the experiment (**Fig. 1**). Questions were projected onto the screen inside the MRI scanner and participants gave answers using a button box with their right hand.

Participants were asked to rate their (hedonic) valence, (emotional) arousal, fear and US expectancy on viewing images of the CS+ and CS- on a nine-step Likert scale from "very unpleasant" to "very pleasant", "quiet and relaxed" to "very excited", "not afraid" to "very afraid" and "US not expected" to "US surely expected", respectively. Post fear acquisition training participants were asked to rate US unpleasantness on a Likert scale from 1 ("not unpleasant") to 9 ("very unpleasant"), and to estimate mean probability (%) that a US occurred after the CS presentation (US expectancy).

Ratings were analyzed using non-parametric ANOVA type statistics for repeated measures with the respective rating as dependent variable and stimulus and time (prior to, post fear acquisition training, post extinction training and post recall) as within-subjects factors as well as their interactions.

2.6. MRI acquisition

All MR images were acquired with the participants lying head first supine inside a whole-body MRI scanner operating at 7T (MAGNETOM 7T, Siemens Healthcare GmbH, Erlangen, Germany) equipped with a 1-channel transmit 32-channel receive array head coil (Nova Medical, Wilmington, MA). To homogenize the radio frequency excitation field (B1), three dielectric pads filled with high-permittivity fluid were placed below and on either side of each participant's upper neck (Teeuwisse et al., 2012). As needed, further cushions were used to fix the head position within the coil.

Functional MRI acquisition was performed using a T2*-weighted CAIPIRINHA-accelerated 3-dimensional echo planar imaging (3D-EPI) sequence (Breuer et al., 2005) with an isotropic resolution of 1.1 mm, a TR/TE of 39 ms/20 ms, and an effective TR of 2340 ms. Covering the cerebellum, the field of view was a coronal, slightly angulated slab with the dimensions 228 × 228 × 88 mm^3. Further imaging parameters were selected as follows: GRAPPA acceleration, 8 × 1, CAIPI shift kz/ky, 0/4, phase and slice partial Fourier factor, 6/8, flip angle, 12°, bandwidth, 1092 Hz/Px, acquisition matrix, 208 × 208 × 80. To facilitate normalization to MNI space a total of five whole brain 3D-EPI volumes (128 slices, effective TR 3744 ms) were acquired immediately before the start of the fMRI acquisition on both days with otherwise identical orientation and parameter settings as the actual fMRI sequence.

2.7. Image processing

3D-EPI were reconstructed offline using MATLAB. All image and fMRI analysis were performed using SPM12 (Wellcome Department of Cognitive Neurology, London, UK) on a platform running MATLAB on a 64-bit Linux machine.

For each session the five EPI volumes acquired for the whole brain were aligned to the first volume and a mean image was calculated. Functional slab volumes covering the cerebellum were aligned to the first volume of the session and co-registered to the respective mean whole brain EPI image.

The high isotropic resolution of 1.1 mm and sufficient white/gray matter contrast allowed for direct segmentation of the mean whole brain EPI volume using the SPM "segment" function. The deformation field yielded by the segmentation was applied to normalize the aligned and coregistered slab volumes into MNI space. No further alignment in between sessions was applied. For artifact removal the Artifact Removal Toolbox (ART) for SPM (version 2015–10, RRID SCR_005994) was applied on functional volumes and movement parameters gained from realignment, generating individual sets of artifact regressors. Finally, functional volumes were smoothed by an isotropic kernel of 3.3 mm.

2.8. fMRI analysis

The first-level analysis was modeled as an event related-design for the entire experiment. Onsets of presentations of the CS+, CS-, US (including the corresponding point in time for unpaired trials, further referred to as no-US), and the onset and end of context presentation were modeled as individual events. Event durations were set to 0 s. Onset and end of context presentation were modeled as regressors of no interest. Individual events were blocked as shown in **Fig. 1**. The experimental phases were split into an early and a late block. In addition, recall was split in blocks of trials with presentations within the acquisition and extinction context. Movement parameters from volume realignment and individual artifact regressors were included as regressors of no interest. During fear acquisition training, events related to the first CS+ trial and the first CS- trial were modeled individually as regressors of no interest since learning could not have started yet.

Our main interest was on context-related cerebellar fMRI signals during recall. We tested the hypotheses whether cerebellar fMRI signals were most prominent related to the CS+ in the acquisition context during recall on day 2. First level main effect contrasts against baseline and appropriate differential first level contrasts were generated and tested in second level *t*-tests. Threshold-free cluster enhancement (TFCE) was applied using the TFCE toolbox for SPM12 (R174, http://dbm.neuro.uni-jena.de/tfce/). To display results, cerebellar (SUIT space) activation maps were plotted on cerebellar flatmaps (Diedrichsen and Zotow, 2015) using TFCE and familywise error (FWE) correction (p<0.05). Activation maps were masked using the SUIT atlas volume (Cerebellum-SUIT.nii) with the inner-cerebellar white matter manually filled in, and maps were projected onto the SUIT atlas volume (Cerebellum-SUIT.nii, Diedrichsen, 2006) to acquire anatomical region labels.

Additionally, mean β values were extracted for two volumes of interest (VOIs) from first level β maps against rest in each of the participants. Vermis was chosen as one of the VOIs, because animal data provide strong evidence that the vermis is involved in fear conditioning (Apps and Strata, 2015). Bilateral lobule VI and Crus I were chosen as second VOI, because the posterolateral cerebellum shows most consistent activation in human fMRI and PET fear conditioning studies (Lange et al., 2015). Non-parametric ANOVA type statistics were performed with mean β values as dependent variable and stimulus, block (early, late) and context as within-subjects factors as well as their interactions.

Finally, a second separate first level analysis was performed on the preprocessed and smoothed functional data. For each experimental phase, all events for each event type (CS, US, no-US, context onset, and context termination) were grouped irrespective of CS trial type (CS+/CS-). Trial-by-trial parameters derived from the individual learning model were applied as parametric modulations, i.e., the mean prediction parameters for the CS events and the mean absolute prediction error parameters for US and no-US events (see **Figs. S2** and **S3**). Again, movement parameters from volume realignment and individual artifact regressors were included as regressors of no interest. Parametric modulation contrasts were tested in second level *t*-tests.

2.9. Modeling

To further confirm the contribution of the cerebellum to renewal effects, predictions of US occurrence were analyzed using a computational model, which does not separate CS+ and CS- from context information *a priori* (Walther et al., 2021). Within the framework of reinforcement learning (Sutton and Barto, 2018), a deep neural network (DNN) was trained to predict the likelihood of a shock for a given visual input, which includes CS and context information. To do so, an artificial agent consisting of a DNN was subjected to a virtual version of the experiment and learned to predict the probability of shock occurring when certain CS were presented. In a second step, the model hyper-parameters were fit to SCRs recorded in the fMRI experiment. The SCRs served as a readout of the participants' state of learning. For each of the experimental phases, the resulting model-derived predictions for the likelihood of shock and the prediction error were then tested in a second-level fMRI evaluation on the single trial contrasts. In each phase, all CS+ event contrasts were setup in a within-subjects ANOVA design, and model prediction values for the likelihood of shock and the prediction error were applied as covariates.

2.9.1. Model and task setup

We simplified the stimuli presented to the agent as RGB line images to improve the agent's ability to generalize between stimuli (**Fig. 2A**). Context change was signaled by reducing the values in one of the color channels. If a CS was paired with the US, the reinforcement was coded as 1, if a CS was not paired with the US, the reinforcement was coded as 0. Like the participants in the study, different virtual agents were subjected to one of the different sequences of pseudo-randomized trials.

2.9.2. Model and training algorithm

The artificial agent consisted of a DNN (LeCun et al., 2015) which was used to represent the mapping between CS and US probability, and a memory module which stored stimulus-reinforcement pairs of the experienced trials (**Fig. 2B**). We call the mapping represented by the DNN with weights θ, the agent's value function $V(s;\theta)$ (Sutton and Barto, 2018). The DNN consisted of two hidden fully connected layers, each of which had 64 units, and an output layer with a single unit, which conveyed the US probability (i.e., the state value). The aforementioned line images were flattened and served as the inputs.

Ten different agents (model instances) with randomly initialized θ were trained on each of the five different trial sequences as follows. On each trial t, the agent first made a prediction about the occurrence of the US v_t given the current stimulus s_t. A reinforcement signal r_t was then given and a US prediction error $\delta_t = r_t - v_t$ computed. The experienced trials were stored in the memory module as experience tuples in the form $e_t = (s_t, r_t, \delta_t)$. Additionally, v_t and δ_t were stored separately for the hyper-parameter fitting procedure described in the next section. At the end of each trial, the agent's value function was adjusted with experience replay (Lin, 1992) as follows: a replay batch B of b experiences was retrieved from the memory module and backpropagation (Rumelhart et al., 1986) was used to update the network weights θ to minimize the mean squared prediction error on B:

$$L(B;\theta) = \frac{1}{b}\sum_{k=1}^{b}\left(r_k - V\left(s_k;\theta\right)\right)^2$$

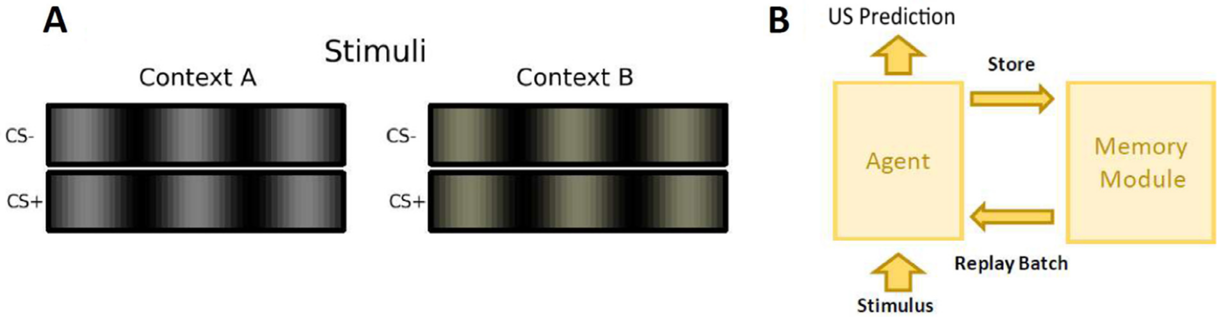


Fig. 2. Modeling. (A) Line images used as input. (B) The computational model used.

where r_k and s_k are the reinforcement signal and the stimulus of replayed experience k, respectively. Experiences were sampled randomly with a probability that was proportional to priority scores

$p = |\delta|\lambda^{\tau}$where τ is the time passed since the experience and λ is a decay factor. So, the priority score depended on the experience's recency and the US prediction error (Schaul et al., 2015). To control the degree of learning from the batch B, we either increased or decreased the number of training repeats (i.e., backpropagation updates) i.

2.9.3. Model and hyper-parameter fitting

The model described above can produce a variety of learning dynamics depending on the choice of the hyper-parameters: replay batch size b, decay factor λ and training repeats i. For instance, a decay factor close to 1 will cause the agent to retain previously learned associations for long durations, which results in slow extinction. Most hyper-parameter combinations result in learning curves that do not resemble participants' learning curves. We therefore fit the model hyper-parameters to the SCRs, which served as a proxy for individual learning rates within participants. Two participants were excluded from the fit due to lacking SCRs.

Between the different trial sequences, the stimuli presented at a given trial were not necessarily the same. Hence, we averaged SCRs from CS+ and CS- trials separately for each trial sequence. We defined the averaged SCRs accordingly as $\bar{Y} = (\bar{y}_{+1}, \dots, \bar{y}_{+N}, \bar{y}_{-1}, \dots, \bar{y}_{-N})$, where $\bar{y}_{+n}$ and $\bar{y}_{-n}$ are the averaged SCRs for the n^{th} CS+ and n^{th} CS- presentations across all participants who completed a given trial sequence, respectively. Analogously, the averaged US predictions of the model were defined as $\bar{V}(b, \lambda, i) = (\bar{v}_{+1}, \dots, \bar{v}_{+N}, \bar{v}_{-1}, \dots, \bar{v}_{-N})$, where $\bar{v}_{+n}$ and $\bar{v}_{-n}$ are the averaged US predictions for the n^{th} CS+ and n^{th} CS- presentations across all model instances, respectively. The quality of the fit of the model was defined as

$$F(b, \lambda, i) = -\sum_{l=1}^{L} w_l E_l(b, \lambda, i)$$

where $E_l(b, \lambda, i)$ is the error between $\bar{Y}_l$ and $\bar{V}_l(b, \lambda, i)$ for trial sequence l

$$E_l(b, \lambda, i) = \|\bar{Y}_l - \bar{V}_l(b, \lambda, i)\| + P_{Acq} + P_{Gen} + P_{Ext} + P_{Ren} + P_{Ret}$$

weighted by the number of participants w_l. Since the error term weights every trial equally, but large changes in behavior occur in a very small number of trials, parameter optimization based only on this error term would miss the rare, but large changes at the transition between experimental phases. To ensure that the overall learning curve in the model resembled that of the participants, we added the following penalty terms, if the model failed to

- acquire the conditioned response by the end of fear acquisition training

$$P_{Acq} = \begin{cases} 0, \ \bar{v}_{Acq,End} \geq 0.5 \\ 10, \ \bar{v}_{Acq,End} < 0.5 \end{cases},$$

where $\bar{v}_{Acq,End}$ is the average US prediction over the last 4 CS+ presentations of fear acquisition training,

- generalize conditioned responses to extinction training

$$P_{Gen} = \begin{cases} 0, \ \bar{v}_{Ext,Start} \geq 0.2 \\ 10, \ \bar{v}_{Ext,Start} < 0.2 \end{cases}$$

where $\bar{v}_{Ext,Start}$ is the average US prediction over the first 4 CS+ presentations of extinction training,

- exhibit renewal of the conditioned response

$$P_{Ren} = \begin{cases} 0, \ \bar{v}_{Rec,Start} \geq 0.2 \ \wedge \ \bar{v}_{Ext,End} \leq 0.05 \\ 10, \ \bar{v}_{Rec,Start} \langle 0.2 \ \vee \ \bar{v}_{Ext,End} \rangle 0.05 \end{cases}$$

where $\bar{v}_{Rec,Start}$ is the average US prediction over the first two CS+ presentations of recall,

- extinguish the conditioned response following renewal

$$P_{Ret} = \begin{cases} 0, \ \bar{v}_{Rec,End} \leq 0.05 \\ 10, \ \bar{v}_{Rec,End} > 0.05 \end{cases}$$

where $\bar{v}_{Rec,End}$ is the average US prediction over the last 4 CS+ presentations of recall.

Table 1
Modeling. Parameter values the grid search was run for.

Grid Search Parameters	
b (batch size)	(Llerena et al., 2002 64, 96, 128)
λ (decay factor)	{0.05, 0.4, 0.5, 0.6, 0.7, 0.8, 0.825, 0.85, 0.895, 1}
i (training repeats)	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

A grid search over the hyper-parameter sets shown in **Table 1** was performed and the fit for each hyper-parameter combination was computed. The model with the best fit was then chosen to generate predictions that were then used as parametric modulations in the fMRI data analysis (**Fig. 3B, C**, see also **Figs. S2, S3**).

3. Results

3.1. Behavioral data

3.1.1. Skin conductance responses (SCRs)

To allow direct comparison with modeling data, SCRs were first analyzed on a trial-by-trial basis.

Habituation (day 1): SCRs related to the CS+ and CS- did not significantly differ (**Fig. 3A**). SCRs to both stimuli in the first trial were significantly higher than in trials 2–4 (least square means test, $p < 0.005$). Non-parametric ANOVA-type statistics revealed a significant main effect of Trial. No significant main effect of Stimulus or Stimulus × Trial interaction were found (**Table 2**).

Fear acquisition training (day 1): SCRs related to the CS+ were significantly higher than to the CS- (**Fig. 3A**, see also **Fig. S1**). Non-parametric

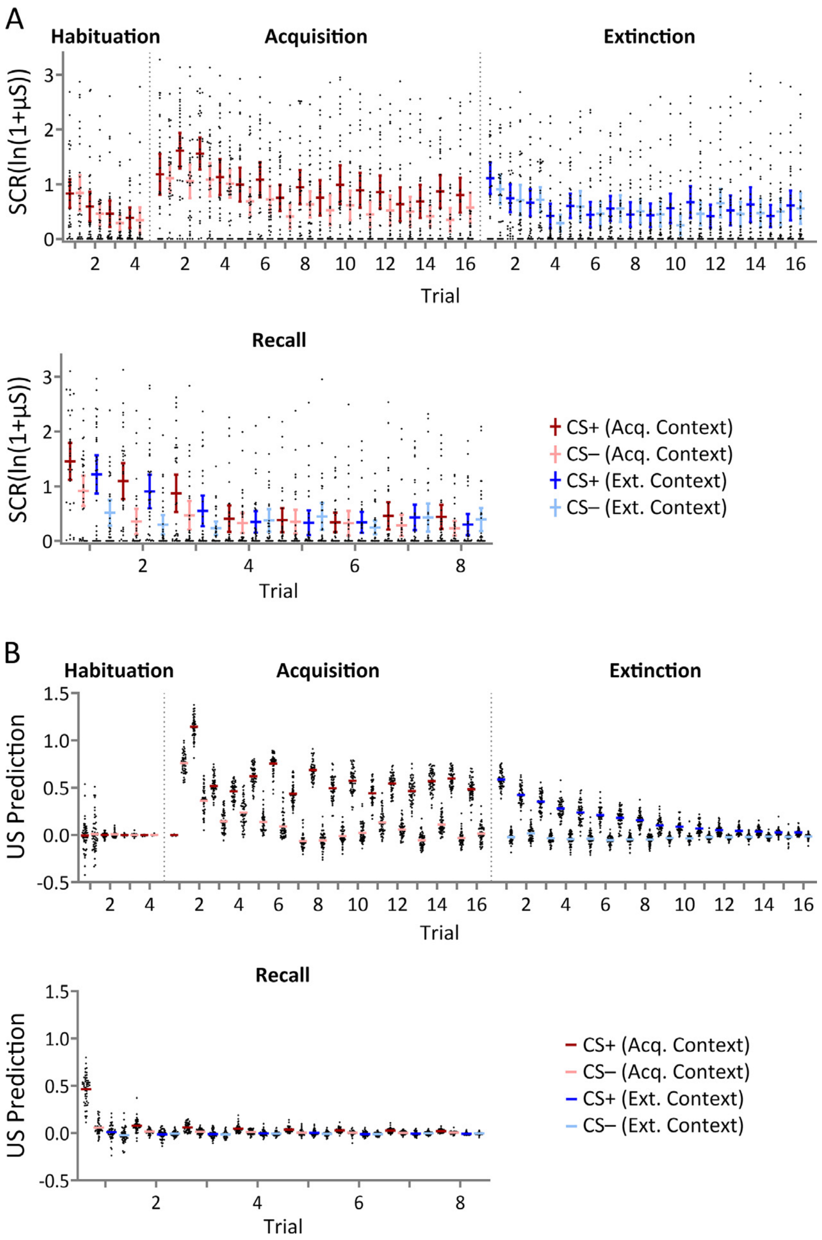


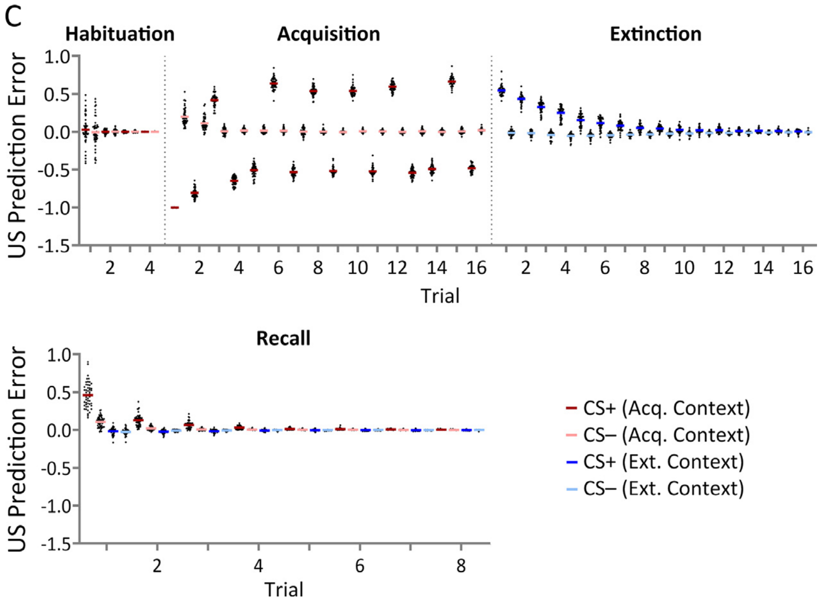
Fig. 3. Fear conditioning and model US prediction and prediction error data. (A) Mean SCRs and individual data on day 1 (habituation, fear acquisition (acq.) and extinction (ext.) training) and day 2 (recall) across trials. Horizontal lines represent group mean values. Vertical lines indicate 95% confidence intervals. Black dots represent individual data points. (B) Mean US predictions and (C) US prediction errors (colored lines) and data of separate model instances (black dots) for habituation, fear acquisition and extinction training and recall across trials. Dark colors = CS+, light colors = CS. Trials in acquisition context are shown in red, trials in extinction context are shown in blue. Note that trial number refers to the number of trials per event type in each phase (for example, 16 CS+ in acq. context for fear acquisition training, 16 CS+ in ext. context for extinction training). Order of the four trial types was pseudorandomized as outlined in the methods (see also Figs S1–S3).

ANOVA-type statistics revealed a significant main effect of Trial and Stimulus. No significant Stimulus × Trial interaction occurred (Table 2).

Extinction training (day 1): SCRs related to the CS+ and CS- did not differ significantly (Fig. 3**A**). Non-parametric ANOVA-type statistics revealed a significant main effect of Trial. No significant main effect of Stimulus or Stimulus × Trial interaction were found (Table 2).

Recall (day 2): SCRs related to the CS+ were significantly higher compared to the CS- and higher in the acquisition context compared to the extinction context (Fig. 3**A**). Additionally, SCRs to the CS+ in the acquisition context were significantly higher than SCRs to the CS+ in the extinction context (least square means test, $p = 0.042$) as well as to the CS- in both the acquisition and extinction context (least square means test, $p < 0.001$). Non-parametric ANOVA-type statistics revealed a significant main effect of Trial, Stimulus, Context, Stimulus × Trial, Stimulus × Context and Trial × Context interactions. No significant Stimulus × Trial × Context interaction emerged (Table 2).

Next, for further illustration of acquisition, extinction, and renewal effects and to allow direct comparison with the main fMRI analysis, SCR data were plotted and re-analyzed comparing the first (early) and second (late) half of each phase (**Fig. S4**). Again, there was no significant differences of SRCs comparing stimuli (CS+ vs. CS-) in the habituation phase, but SCRs were significantly higher related to the CS+ compared to the CS- during fear acquisition training showing that participants had learned that CS+ presentation was followed by electric shock. As indicated in the methods, participants were instructed that should they perceive a pattern between CS and US presentations, the experimenter would not change it during the experiment. As in previous studies, semi-instructed fear conditioning likely prevented a significant Stimulus x Trial interaction (see e.g., Ernst et al. 2019). Whereas SCRs to the CS+ were significantly higher compared to the CS- during fear acquisition training, no difference was seen during extinction training, indicating that extinction of the learned fear association had occurred. In early recall, SCRs were significantly higher in CS+ compared to CS- trials, and in the acquisition compared to the extinction context, indicating context-dependent renewal effects. Statistical findings are summarized in **Table S1**.



3.1.2. Questionnaires

Valence, arousal, and fear ratings. Prior to fear acquisition training, valence, arousal, and fear ratings of the CS+ and CS- were not significantly different from each other (Table 3). Post fear acquisition training, the CS+ was rated as less pleasant, higher arousing and more fearful compared to the CS-. These differences remained until the end of recall. Non-parametric ANOVA-type statistic revealed a significant main effect of Time, Stimulus, and a Stimulus × Time interaction (Table 3). *Post-hoc* tests showed significant differences between stimuli post fear acquisition training, extinction training and recall (least square means tests, Valence: all p values < 0.001; Arousal: all p values < 0.003; Fear: all p values < 0.0052), but not prior to fear acquisition training (least square means test, all $p > 0.989$).

US unpleasantness, CS-US contingency, and US expectancy. Median US unpleasantness was rated 7 (IQR 5.25–8) post fear acquisition training. Post fear acquisition training, participants reported that they recognized a pattern between CS+ and US after 3.03 ± 2.96 min or 2.74 ± 2.05 electric shocks. Across all participants, mean probability that a US occurred after CS+ presentation was estimated as 67.7 ± 16.87%, and after CS- presentation as 1.29 ± 4.28% (0% probability by 28 out of 31 (90%) participants). Prior to fear acquisition training, reported US expectation after CS+ and CS- were not significantly different from each other (Table 3). Post fear acquisition training, participants reported a higher US expectation after CS+ compared to the CS- and this difference remained until the end of recall. Non-parametric ANOVA-type statistics revealed a significant main effect of Time, Stimulus, and a Stimulus × Time interaction (Table 3). *Post-hoc* tests showed significant differences between stimuli post fear acquisition training, extinction training and recall (least square means tests, all p values < 0.002), but not prior to fear acquisition training (least square means test, $p = 1.0$).

3.2. fMRI data

In none of the experimental phases cerebellar fMRI activations were significantly different comparing CS+ and CS-, with the only exception of early recall in the acquisition context. Lack of differences was expected during habituation, during late extinction training and late recall. Lack of differences comparing CS+ and CS- were unexpected during fear acquisition training. Because no significant differential activations emerged during fear acquisition training, the decision was made to report data of CS+ and CS- activations against baseline in all experimental phases. Possible reasons for the lack of differential cerebellar activations during fear acquisition training will be addressed in the discussion section.

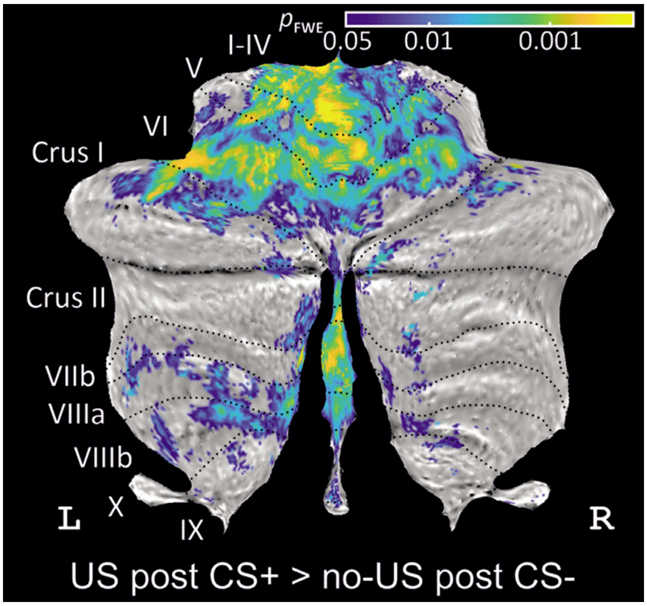


Fig. 4. Cerebellar activation related to the presentation of the aversive US [contrast 'US post CS+ > no US post CS-' during fear acquisition training]. Cerebellar activations in SUIT space projected on a cerebellar flatmap (Diedrichsen and Zotow, 2015). All contrasts collapsed over early and late fear acquisition blocks and calculated using TFCE and FWE correction ($p < 0.05$). CS = conditioned stimulus; L = left; R = right; SUIT = spatially unbiased atlas template of the cerebellum; TFCE = threshold-free cluster-enhancement; FWE = family-wise error rate; US = unconditioned stimulus.

3.3. Cerebellar activation related to presentation of the aversive stimulus (US)

Cerebellar activation related to presentation of the aversive stimulus [contrast 'US post CS+ > no-US post CS-'] was observed within the cerebellar vermis and both cerebellar hemispheres (Fig. 4; see also Table 4). Most prominent activations were found in the anterior and posterior

Table 2
Skin conductance responses. Results of the non-parametric ANOVA-type statistics for repeated measures for habituation, fear acquisition training, extinction training and recall.

Factor	Df[†]	*F*	*p*
Skin conductance responses			
Habituation			
Stimulus	1	2.92	0.098
Trial	2.45	9.93	**<0.001**[***]
Stimulus × Trial	2.23	0.23	0.821
Fear acquisition training			
Stimulus	1	18.26	**<0.001**[***]
Trial	8.35	10.88	**<0.001**[***]
Stimulus × Trial	9.01	1.25	0.260
Extinction training			
Stimulus	1	1.05	0.305
Trial	8.2	4.16	**<0.001**[***]
Stimulus × Trial	9.22	1.10	0.362
Recall			
Stimulus	1	24.48	**<0.001**[***]
Trial	4	12.37	**<0.001**[***]
Context	1	4.20	**0.041**[*]
Stimulus × Trial	4.67	7.36	**<0.001**[***]
Stimulus × Context	1	3.88	**0.049**[*]
Trial × Context	5.42	2.26	**0.041**[*]
Stimulus × Trial × Context	5.61	0.60	0.722
Valence			
Stimulus	1	78.95	**<0.001**[***]
Time	3.25	6.88	**<0.001**[***]
Stimulus × Time	3.18	16.46	**<0.001**[***]
Arousal			
Stimulus	1	89.58	**<0.001**[***]
Time	3.5	11.03	**<0.001**[***]
Stimulus × Time	3.51	8.46	**<0.001**[***]
Fear			
Stimulus	1	72.66	**<0.001**[***]
Time	2.95	8.44	**<0.001**[***]
Stimulus × Time	3.52	12.74	**<0.001**[***]
US expectancy			
Stimulus	1	155.46	**<0.001**[***]
Time	2.45	11.65	**<0.001**[***]
Stimulus × Time	3.39	21.03	**<0.001**[***]

[*] Significant results at $p < 0.05$.

[***] Significant results at $p < 0.001$.

[†] Since, in general, ranked observations are heteroscedastic (Akritas, 1990), assumption of an arbitrary covariance matrix (Brunner et al., 2002) is suggested and therefore, degrees of freedom are appropriately adjusted (Noguchi et al., 2012).

vermis (local maxima in vermal lobules I-IV, V, VIII) and the left hemisphere (ipsilateral to the presentation of the US; local maxima in lobules V, VI).

3.4. Cerebellar activation related to CS+ and CS- presentation during fear acquisition training

During early fear acquisition training, cerebellar activation related to CS+ presentation (contrast ‘CS+ > rest’) was observed within the left cerebellar hemisphere (**Table 4**), with local maxima in lobules VI and Crus I. No voxels were significantly activated related to CS- presentation or when comparing stimulus type events (contrast ‘CS+ > CS-’ and ‘CS- > CS+’).

During late fear acquisition training, CS+ and CS- related cerebellar activations (contrast ‘CS+ > rest’ and ‘CS- > rest’) were found in the vermis (local maximum in lobule VI) and both hemispheres (local maxima in lobules VI and Crus I) (**Fig. 5**; see also **Table 4**). No voxels were significantly activated for the comparison of stimulus type events (contrast ‘CS+ > CS-’ and ‘CS- > CS+’).

Second level *t*-tests did not reveal activations at the time the US was expected and did not occur (using an uncorrected threshold of $p < 0.001$). This lack of significant cerebellar activations is at variance with previous findings of our group (Ernst et al., 2019). This is likely explained by differences in the complexity of the fear conditioning paradigm, which will be addressed in detail in the discussion.

3.5. Cerebellar activation related to the CS+ and CS- presentation during extinction training

During early extinction training, CS+ and CS- related cerebellar activation (contrast ‘CS+ > rest’ and ‘CS- > rest’) was observed within both cerebellar hemispheres (**Fig. 6**; see also **Table 5**), with local maxima in lobules VI and Crus I. No voxels were significantly activated for the

Table 3
Fear conditioning questionnaires. Median (interquartile range) valence, arousal, fear and US expectancy ratings prior to fear acquisition training, post fear acquisition training, post extinction training and post recall.

	Time of assessment			
Stimulus	Prior acquisition	Post acquisition	Post extinction	Post recall
		Valence ratings (1 – uncomfortable, 9 – comfortable)		
CS+	6 (7–5)	**3 (4–2)***,†	**5 (6–4)**†	**5 (6–4)**†
CS-	6 (7–5)	**8 (9–7)***,†	**7 (8–6)**†	**8 (9–7)***,†
		Arousal ratings (1 – very calm, 9 – very nervous)		
CS+	4 (6–2)	**7 (8–6)***,†	**4 (6–2)**†	**4 (6–2)**†
CS-	3 (4–2)	**2 (3–1)**†	**2 (3–1)**†	**1 (2–1)***,†
		Fear ratings (1 – not afraid, 9 – very afraid)		
CS+	1 (2–1)	**6 (7–4)***,†	**3 (5–2)***,†	**3 (5–2)***,†
CS-	1 (2–1)	**1 (3–1)**†	**1 (2–1)**†	**1 (2–1)**†
		US expectancy ratings (1 – US not expected, 9 – US surely expected)		
CS+	1 (5–1)	**7 (8–7)***,†	**5 (6–2)***,†	**5 (6–3)***,†
CS-	1 (4–1)	**2 (3–1)**†	**2 (3–1)** †	**1 (2–1)**†

Statistically significant differences are shown in bold (least square means tests, $p < 0.05$):.

* significant differences between prior to and post fear acquisition training.

† significant differences between CS+ and CS-.

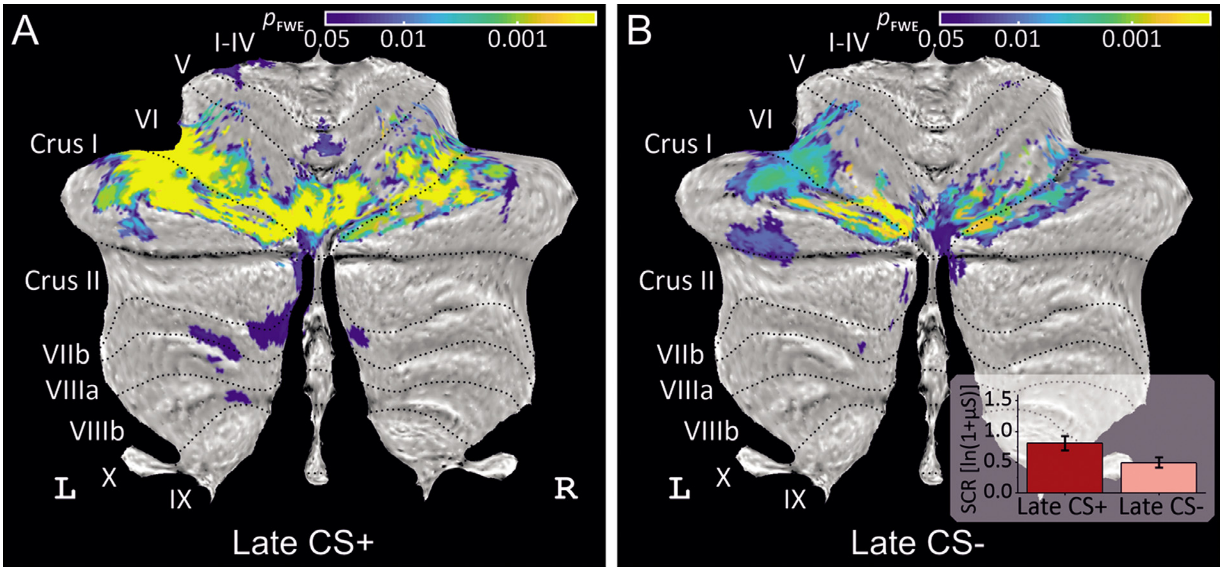


Fig. 5. Cerebellar activation related to the CS+ and CS- during late fear acquisition training. Cerebellar activations during the presentation of (**A**) CS+ in the late fear acquisition block and (**B**) CS- in the late fear acquisition block in SUIT space projected on a cerebellar flatmap (Diedrichsen and Zotow, 2015). All contrasts are calculated using TFCE and FWE correction ($p < 0.05$). Insert shows mean skin conductance responses (SCRs) during late fear acquisition training. Error bars indicate standard errors (SE). CS = conditioned stimulus; L = left; R = right; SUIT = spatially unbiased atlas template of the cerebellum; TFCE = threshold-free cluster-enhancement; FWE = family-wise error rate.

comparison of stimulus type events (contrast ‘CS+ > CS-’ and ‘CS- > CS+’).

During late extinction training, CS+ presentation related cerebellar activation (contrast ‘CS+ > rest’) was also observed in both cerebellar hemispheres (**Fig. 6**; see also **Table 5**), with local maxima in lobules VI and Crus I. During CS- presentation (contrast ‘CS+ > rest’), a small cluster of voxels was active in left Crus I. No voxels were significantly activated for the comparison of stimulus type events (contrast ‘CS+ > CS-’ and ‘CS- > CS+’).

3.6. Cerebellar activation related to the CS+ and CS- presentation during recall

Acquisition context. During early recall, CS+ related cerebellar activation in the acquisition context (contrast ‘CS+ > rest’) was observed within the cerebellar vermis (local maximum in lobule VI) and both cerebellar hemispheres (local maxima in lobules VI and Crus I; **Fig. 7**; see also **Table 6**). During CS- presentation in the acquisition context (contrast ‘CS- > rest’), a small cluster of voxels was active in left lobule VI and Crus I. The comparison of stimulus types (contrast ‘CS+ > CS-’) revealed three small clusters in right lobules I-IV and VI during early recall (**Table 6**). The contrast ‘CS- > CS+’ revealed no significant differences.

During late recall, CS+ related cerebellar activation in the acquisition context (contrast ‘CS+ > rest’) was observed within the left cerebellar hemisphere (**Table 6**), with local maxima in lobules VI and Crus I. No voxels were significantly activated during CS- (contrast ‘CS- > rest’) presentation or comparing stimulus types (contrast ‘CS+ > CS-’ and ‘CS- > CS+’) in the acquisition context.

Extinction context. During early recall, CS+ related cerebellar activation in the extinction context (contrast ‘CS+ > rest’) was less compared to the acquisition context. Some remaining cerebellar activations were

Table 4
Fear acquisition training. Activation clusters are reported which were significant after application of threshold-free cluster-enhancement (TFCE) at $p < 0.05$ FWE corrected level (t-tests). Displayed are all clusters of ≥ 20 mm^3. In each cluster, up to three maxima are listed separated by ≥ 8 mm. ncl. = nucleus.

Index	Location (lobule)	Side	SUIT coordinates/mm			Cluster size/mm^3	p_{FWE}	TFCE
US post CS+ > no US post CS-								
1	Extended cluster	left VI (8023), right VI (4999), right V (4235), left V (4230), left I-IV (3387), right I-IV (3282), left Crus I (2688), left VIIIb (1505), vermal VI (1454), left VIIIa (1204), gray matter (1133), vermal VIIIa (889), right Crus I (685), left VIIb (609), right VIIIb (549), right IX (371), vermal VIIIb (324), left IX (247), right VIIb (244), vermal VIIb (243), right Crus II (231), left Crus II (218), right VIIIa (203), vermal IX (194), vermal Crus II (190), left dentate ncl. (138), right interposed ncl. (68), right dentate ncl. (64), left interposed ncl. (64), vermal X (7), left fastigial ncl. (7), right fastigial ncl. (2), vermal Crus I (1)						
	I-IV	left	−13	−41	−22	41,688	<0.001	6189.3
	I-IV	left	−2	−52	−23		<0.001	6062.1
	I-IV	left	−22	−59	−19		<0.001	5935.4
2	gray matter		15	−37	−43	76	0.012	2055.8
4	Crus I	left	−46	−43	−31	16	0.033	1625.8
CS+: early fear acquisition training								
1	VI	left	−34	−51	−31	68	0.026	877.5
CS-: early fear acquisition training								
	no significant voxels							
CS+: late fear acquisition training								
1	Extended cluster	left VI (3322), left Crus I (2710), right VI (1756), vermal VI (804), right Crus I (716), gray matter (469), right V (259), left V (111), right I-IV (32), left dentate ncl. (27), vermal Crus II (14), left I-IV (12), left Crus II (10), vermal Crus I (3)						
	VI	left	−7	−77	−18	10,245	<0.001	3297.7
	VI	left	−28	−72	−20		<0.001	3157.5
	Crus I	left	−40	−57	−29		<0.001	2991.9
2	Extended cluster	gray matter (265), left I-IV (19)						
	gray matter		−12	−35	−31	284	0.016	991.9
	gray matter		−8	−44	−29		0.019	956.7
3	Extended cluster	left Crus II (226), left VIIb (62), gray matter (58), left dentate ncl. (19)						
	Crus II	left	−5	−78	−36	365	0.02	947
	VIIb	left	−12	−72	−41		0.038	844.7
4	Extended cluster	left VIIb (161), left VIIIa (64), gray matter (1)						
	VIIb	left	−28	−66	−51	226	0.026	917.9
	VIIIa	left	−18	−67	−46		0.035	856.4
5	Extended cluster	right VIIb (63), right Crus II (36)						
	Crus II	right	6	−76	−42	99	0.038	840.8
	VIIb	right	9	−69	−38		0.048	797.1
6	Crus I	right	42	−55	−35	55	0.041	831.2
7	VIIIb	left	−13	−61	−50	42	0.041	827.8
CS-: late fear acquisition training								
1	Extended cluster	left VI (1781), left Crus I (1722), right VI (1021), right Crus I (522), vermal VI (353), gray matter (272), left V (17), left Crus II (15), vermal Crus II (2), right Crus II (2)						
	VI	left	−10	−82	−20	5707	<0.001	2048.8
	VI	left	−18	−79	−20		<0.001	1818.8
	VI	left	−29	−71	−20		<0.001	1717.2
2	Extended cluster	gray matter (169), left VI (2)						
	gray matter		−17	−48	−30	171	0.024	925.4
	gray matter		−10	−51	−24		0.042	812.5
3	gray matter		−9	−71	−33	39	0.042	809
CS+ > CS-								
	no significant voxels							
CS+ > CS-: early fear acquisition training								
	no significant voxels							
CS+ > CS-: late fear acquisition training								
	no significant voxels							

observed within the cerebellar vermis (local maximum in lobule VI) and both cerebellar hemispheres (local maxima in lobules VI and Crus I in both hemispheres; lobules I-IV and V in the right hemisphere; Fig. 7; see also Table 6). During CS- presentation in the extinction context (contrast 'CS- > rest'), cerebellar activation was observed within the cerebellar vermis (local maximum in lobule VI) and both cerebellar hemispheres (local maximum in lobules V-VI and Crus I; Fig. 7; see also Table 6).

During late recall, CS+ related cerebellar activation in the extinction context (contrast 'CS+ > rest') was observed only within a small cluster of voxels in the left cerebellar hemisphere (Table 6), with local maxima in lobules VI and Crus I. No voxels were significantly activated during CS- (contrast 'CS- > rest') presentation or when comparing stimulus type events in the extinction context (contrast 'CS+ > CS-' and 'CS- > CS+').

Comparison of acquisition and extinction context. Close inspection of Fig. 8 showed that cerebellar activation related to presentation of the CS+ was higher in the acquisition context compared to the extinction context during early recall. This was true for areas in the vermis and posterolateral cerebellum. Voxel-wise statistical analysis revealed no voxels with a significant difference comparing CS+ and CS- related activations in the two contexts (contrasts 'CS+ in the acquisition context > CS+ in the extinction context' vs. 'CS- in the acquisition context > CS- in the extinction context').

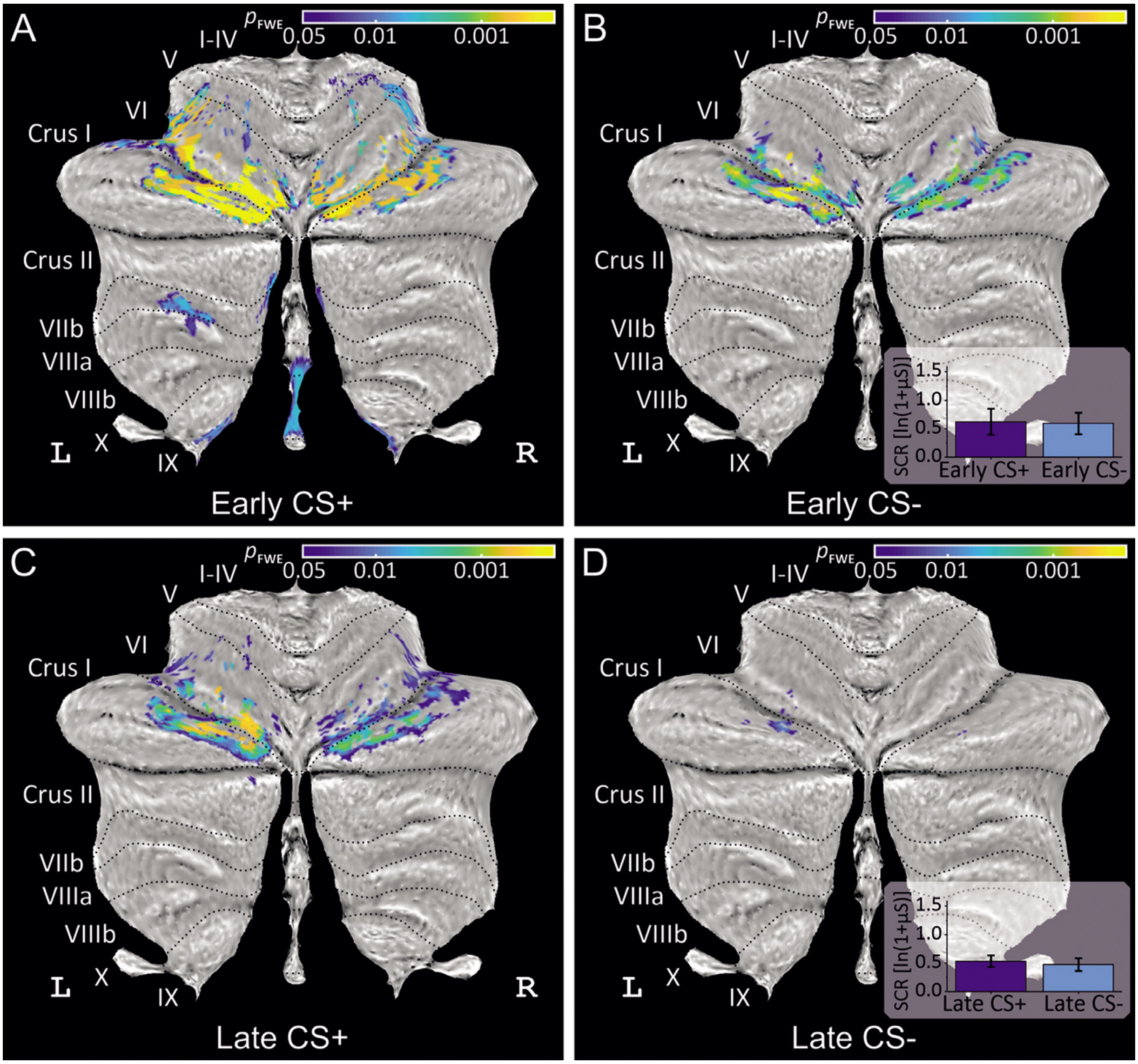


Fig. 6. Cerebellar activation related to the CS+ and CS- during extinction training. Cerebellar activations during the presentation of (**A**) CS+ in the early extinction block, (**B**) CS- in the early extinction block, (**C**) CS+ in the late extinction block and (**D**) CS- in the late extinction block in SUIT space projected on a cerebellar flatmap (Diedrichsen and Zotow, 2015). All contrasts are calculated using TFCE and FWE correction ($p < 0.05$). Insert shows mean skin conductance responses (SCRs) during early and late extinction training. Error bars indicate standard errors (SE); R = right; SUIT = spatially unbiased atlas template of the cerebellum; TFCE = threshold-free cluster-enhancement; FWE = family-wise error rate.

VOI analysis of β values in the vermis and lobule VI/Crus I bilaterally, however, provided some evidence of context-dependent activations during recall (Fig. 8). During recall, non-parametric ANOVA type statistics on mean β values in lobules VI and Crus I revealed a significant Block × Context interaction ($F_1 = 4.27, p = 0.048$). No other significant main effects or interactions were observed (all p values > 0.14). Post hoc exploratory analysis of the Block × Context differences revealed significantly higher mean β values in the acquisition context in early vs. late recall blocks ($p = 0.03$ uncorrected, $p = 0.12$ adjusted for multiple comparisons). During recall, non-parametric ANOVA type statistics on mean β values in vermis did not reveal any significant main effect or interaction (all p values > 0.14). The Stimulus type × Block × Context effect, however, was close to significance ($F_1 = 0.39, p = 0.055$).

There was no significant effect of stimulus type (CS+ or CS-) during habituation (vermis: $F_1 = 0.78, p = 0.38$; VI and Crus I: $F_1 = 0.27, p = 0.61$). During fear acquisition training, mean β values in late trials were significantly higher compared to early trials (vermis: $F_1 = 3.39, p = 0.075$; VI and Crus I: $F_1 = 7.14, p = 0.012$). The CS (CS+ vs. CS-) effect (vermis: $F_1 = 2.76, p = 0.10$; VI and Crus I: $F_1 = 0.01, p = 0.92$) and Stimulus type × Block interaction (vermis: $F_1 = 0.12, p = 0.73$; VI and Crus I: $F_1 = 0.67, p = 0.42$) effects were not significant. During extinction training, non-parametric ANOVA type statistics on mean β values in the vermis did not reveal any significant effect of Stimulus type ($F_1 = 1.96, p = 0.17$), Block ($F_1 = 0.01, p = 0.93$) or Stimulus type × Block interaction ($F_1 = 1.00, p = 0.32$). In lobules VI and Crus I, mean β values in CS+ trials were significantly higher compared to CS- trials ($F_1 = 4.47, p = 0.043$). The Block (early vs late; $F_1 = 1.4, p = 0.25$) and Stimulus type × Block interaction (vermis: $F_1 = 0.16, p = 0.69$) effects were not significant.

3.7. Parametric modulation with model predictions for shock probability and prediction errors

We computed effects of parametric modulation of the learning model-derived predictions on the fMRI signal of CS and no-US events. As described above, CS events refer to the time of CS onset, and paramet-

Table 5
Extinction training. Activation clusters are reported which were significant after application of threshold-free cluster-enhancement (TFCE) at $p < 0.05$ FWE corrected level (t-tests). Displayed are all clusters of ≥ 20 mm^3. In each cluster, up to three maxima are listed separated by ≥ 8 mm. ncl. = nucleus.

Index	Location (lobule)	Side	SUIT coordinates/mm			Cluster size/mm^3	p_{FWE}	TFCE
CS+: early extinction training								
1	Extended cluster	left VI (893), left Crus I (386), vermal VI (83), left V (8)						
	Crus I	left	−36	−77	−23	1370	<0.001	2546.6
	VI	left	−28	−72	−20		<0.001	2520.4
	VI	left	−9	−77	−16		<0.001	2367.7
2	Extended cluster	right VI (568), right Crus I (310), right V (31), right I-IV (15), vermal VI (13), gray matter (2)						
	VI	right	31	−63	−19	939	<0.001	1672.1
	Crus I	right	14	−83	−20		0.001	1554.8
	VI	right	36	−69	−21		0.001	1469.2
3	Extended cluster	right V (70), right VI (29)						
	VI	right	24	−53	−16	99	0.001	1183.9
	V	right	21	−44	−16		0.003	1014.7
4	Extended cluster	vermal IX (363), left dentate ncl. (176), gray matter (151), left IX (146), vermal VIIIb (121), right IX (111), left interposed ncl. (78), right interposed ncl. (47), vermal VIIIa (21), right dentate ncl. (19), right VIIIa (4), vermal X (1)						
	dentate ncl.	left	−7	−63	−32	1238	0.001	1083.6
	VIIIb	vermal	0	−60	−37		0.003	1022.4
	IX	right	8	−55	−33		0.005	965.1
5	Extended cluster	left VIIb (318), left VIIIa (34), gray matter (7)						
	VIIb	left	−28	−63	−47	359	0.004	988.5
	VIIb	left	−35	−63	−54		0.014	832
	VIIIa	left	−27	−66	−57		0.017	809.9
6	Extended cluster	left VI (59), left V (27)						
	V	left	−22	−52	−16	86	0.008	905.4
	VI	left	−21	−62	−15		0.013	837.7
CS-: early extinction training								
1	Extended cluster	left VI (177), left Crus I (137)						
	VI	left	−29	−71	−20	314	<0.001	1791.7
	Crus I	left	−41	−65	−23		<0.001	1553.8
	VI	left	−20	−73	−18		0.007	951.4
2	Extended cluster	right Crus I (178), right VI (153), gray matter (2)						
	Crus I	right	42	−68	−22	333	0.001	1366.5
	VI	right	31	−63	−19		0.001	1341.9
	VI	right	27	−74	−19		0.001	1222.3
3	Extended cluster	left VI (82), left Crus I (23), vermal VI (17)						
	VI	left	−6	−78	−17	122	0.001	1250.2
	Crus I	left	−17	−83	−22		0.003	1090.2
4	VI	right	11	−75	−15	52	0.002	1152.2
CS+: late extinction training								
1	Extended cluster	left Crus I (400), left VI (257), left V (23)						
	VI	left	−28	−72	−20	680	<0.001	1884.9
	Crus I	left	−36	−79	−24		0.001	1696.6
	Crus I	left	−42	−72	−24		0.001	1629.5
2	Extended cluster	right Crus I (241), right VI (156), gray matter (1)						
	Crus I	right	30	−77	−21	398	0.001	1557.7
	Crus I	right	18	−84	−21		0.001	1485.8
	Crus I	right	40	−73	−22		0.002	1451.9
3	Extended cluster	right VI (107), right Crus I (97)						
	Crus I	right	42	−61	−23	204	0.006	1218.5
	VI	right	39	−47	−25		0.019	995
	Crus I	right	50	−60	−25		0.03	901.5
4	Extended cluster	left VI (89), left Crus I (26)						
	VI	left	−39	−51	−25	115	0.014	1047.8
	VI	left	−30	−53	−21		0.02	986.8
CS-: late extinction training								
	no clusters of ≥ 20 mm^3							
CS+ > CS-								
	no significant voxels							
CS+ > CS-: early extinction training								
	no significant voxels							
CS+ > CS-: late extinction training								
	no significant voxels							

ric modulation was based on trial-by-trial model-derived predictions of US/shock probability. No-US events refer to the time the US is expected in reinforced CS+ trials but does not occur in unreinforced CS+ and CS- trials. Therefore, in no-US events, parametric modulation was based on trial-by-trial model-derived prediction errors.

In fear extinction training, parametric modulation effects for CS and prediction values showed small significant clusters in cerebellar lobules Crus I and VI (at an uncorrected threshold of $p < 0.001$; see Fig. 9A, C, and Table 7). No significant clusters were observed in the acquisition phase considering clusters of ≥ 20 mm^3. In both the acquisition and extinction phase, more prominent effects were observed for the parametric modulations of model-derived prediction error values for the omission of US events at CS termination (no-US) (at an uncorrected threshold of $p < 0.001$; Fig. 9B, D, and Table 7). Note that parametric modula-

Table 6

Recall. Activation clusters are reported which were significant after application of threshold-free cluster-enhancement (TFCE) at $p < 0.05$ FWE corrected level (t-tests). Displayed are all clusters of ≥ 20 mm^3. In each cluster, up to three maxima are listed separated by ≥ 8 mm.

Index	Location (lobule)	Side	SUIT coordinates/mm			Cluster size/mm^3	p_{FWE}	TFCE
Acquisition context								
CS+: early recall								
1	Extended cluster	left VI (2090), right VI (1755), left Crus I (1477), right Crus I (807), vermal VI (790), right V (201), left V (110), gray matter (32), right I-IV (31)						
	Crus I	left	−27	−79	−22	7293	<0.001	2221.2
	VI	left	−35	−68	−22		<0.001	2212.1
	Crus I	left	−46	−68	−25		<0.001	2201.7
2	Extended cluster	left Crus I (444), left VI (265), gray matter (25), left Crus II (6)						
	Crus I	left	−39	−50	−32	740	0.005	1169.2
	VI	left	−34	−42	−38		0.019	956.1
	Crus I	left	−43	−41	−40		0.033	866
3	Extended cluster	right VI (58), right V (22)						
	V	right	30	−42	−24	80	0.033	863.8
	VI	right	38	−43	−27		0.035	854.9
5	Crus I	right	44	−47	−40	79	0.037	845.6
4	Crus I	left	−31	−73	−33	36	0.043	817.8
CS-: early recall								
1	Extended cluster	left VI (48)						
	VI	left	−36	−66	−22	48	0.028	926.2
	VI	left	−33	−53	−22		0.036	884.3
CS+: late recall								
	no clusters of ≥ 20 mm^3							
CS-: late recall								
	no significant voxels							
CS+ > CS-								
	no significant voxels							
CS+ > CS-: early recall								
1	I-IV	right	3	−57	−4	48	0.032	833.7
2	I-IV	right	5	−51	−7	37	0.033	825.0
3	VI	right	29	−65	−21	25	0.036	820.1
CS+ > CS-: late recall								
	no significant voxels							
Extinction context								
CS+: early recall								
1	Extended cluster	right VI (605), right Crus I (313), right V (137), right I-IV (48), vermal VI (11), gray matter (1)						
	VI	right	37	−62	−21	1115	0.001	1617.7
	VI	right	33	−45	−24		0.002	1357.9
	Crus I	right	18	−82	−20		0.002	1316.8
2	Extended cluster	left VI (671), left Crus I (643), vermal VI (156), left Crus II (67), gray matter (9)						
	Crus I	left	−27	−79	−22	1546	0.001	1615.4
	VI	left	−18	−79	−20		0.001	1551.9
	VI	left	−33	−69	−21		0.001	1498.5
3	VI	right	15	−73	−15	39	0.02	880.3
	VI	right	8	−78	−17		0.027	850.9
CS-: early recall								
1	Extended cluster	left VI (904), left Crus I (700), left V (276)						
	VI	left	−27	−70	−19	1880	<0.001	1966.2
	VI	left	−26	−54	−19		<0.001	1761.8
	Crus I	left	−34	−73	−22		<0.001	1723.7
2	Extended cluster	right VI (1332), right V (460), right Crus I (226), vermal VI (107), gray matter (12), right I-IV (10)						
	VI	right	29	−71	−19	2147	<0.001	1827.2
	VI	right	12	−75	−16		<0.001	1691.5
	V	right	25	−45	−19		0.001	1430.5
3	Extended cluster	left VI (543), vermal VI (277), left Crus I (150)						
	VI	vermal	−4	−70	−15	970	0.001	1504.1
	VI	left	−9	−78	−18		0.001	1470.4
	Crus I	left	−16	−83	−22		0.001	1467.2
4	V	left	−17	−41	−20	23	0.04	801.5
5	Crus I	left	−51	−58	−29	56	0.044	788
CS+: late recall								
1	Crus I	left	−33	−79	−23	21	0.015	994.3
CS-: late recall								
	no significant voxels							
CS+ > CS-								
	no significant voxels							
CS+ > CS-: early recall								
	no significant voxels							
CS+ > CS-: late recall								
	no significant voxels							

Table 7
Parametric modulation of learning model-derived prediction parameters. Displayed are all clusters of ≥ 20 mm^3. In each cluster, up to three maxima are listed separated by ≥ 8 mm. Thresholded at $p < 0.001$ uncorrected.

Index	Location (lobule)	Side	SUIT coordinates/mm			Cluster size/mm^3	*T*
CS × prediction in fear acquisition training							
	no significant clusters						
CS × prediction in extinction training							
1	Extended cluster	right VI (99), right V (61)					
	VI	right	31	−47	−21	160	4.58
	VI	right	24	−52	−16		4.58
	V	right	22	−44	−16		3.69
2	VI	right	32	−63	−19	40	3.71
3	Crus II	right	7	−78	−42	50	3.71
4	VI	right	39	−43	−26	83	5.62
5	VI	left	−11	−79	−18	45	5.04
6	V	left	−25	−30	−27	27	4.98
7	Extended cluster	left VI (175), left Crus I (44)					
	VI	left	−41	−46	−29	219	4.95
	VI	left	−36	−40	−27		4.31
8	VIIb	left	−27	−68	−56	41	4.35
9	IX	vermal	−1	−56	−38	31	4.07
10	VIIIb	left	−26	−47	−57	24	3.87
11	Crus II	left	−5	−82	−32	23	3.82
12	Crus II	left	−7	−77	−41	21	3.72
CS × prediction in recall							
1	Extended cluster	left VI (109), left Crus I (96)					
	Crus I	left	−35	−47	−35	205	5.20
	VI	left	−31	−58	−32		4.50
2	V	left	−25	−35	−31	42	5.20
3	Crus I	right	38	−50	−33	101	4.52
4	VI	right	8	−72	−14	22	4.50
5	Crus I	right	38	−71	−22	21	4.42
no-US × prediction error in fear acquisition training							
1	Extended cluster	left VI (208), left Crus I (17)					
	VI	left	−20	−73	−25	225	4.71
	VI	left	−22	−67	−31		4.67
2	Crus I	left	−34	−71	−26	72	4.69
3	Crus I	left	−6	−79	−27	77	4.51
4	VI	left	−34	−55	−29	39	4.43
5	Crus I	left	−46	−52	−33	57	4.18
no-US × prediction error in extinction training							
1	Extended cluster	right VI (171), right Crus I (159)					
	Crus I	right	15	−77	−24	330	6.44
	Crus I	right	26	−81	−21		5.07
	Crus I	right	36	−76	−22		4.99
2	Extended cluster	left Crus I (1254), left VI (506), vermal VI (178), left Crus II (79), right VI (55), vermal Crus I (3)					
	VI	left	−7	−81	−21	2075	6.44
	VI	left	−18	−78	−20		6.21
	Crus I	left	−16	−77	−32		5.67
3	Crus I	left	−27	−67	−38	58	6.06
4	Extended cluster	left Crus I (281)					
	Crus I	left	−44	−69	−28	281	5.85
	Crus I	left	−50	−60	−31		4.37
5	Crus I	right	43	−61	−23	23	5.40
6	Crus I	left	−45	−52	−33	72	5.30
7	VIIb	left	−28	−68	−51	66	5.11
8	VI	right	27	−58	−17	31	5.04
9	Extended cluster	left VI (147), left Crus I (39)					
	VI	left	−31	−65	−25	186	4.65
	VI	left	−39	−60	−23		3.64
10	VI	right	38	−42	−26	23	4.56
11	VIIIb	vermal	1	−60	−34	23	4.42
12	Crus I	right	47	−53	−34	92	4.41
13	Crus I	left	−46	−69	−34	40	4.36
14	VI	left	−41	−46	−28	22	4.20
15	Crus II	right	7	−77	−30	31	4.13
16	Crus I	right	47	−64	−25	29	3.91
no-US × prediction error in recall							
1	I-IV	left	−2	−50	−3	38	5.04
2	Crus I	right	26	−75	−25	77	4.96
3	VI	left	−16	−77	−22	24	4.69
4	IX	left	−7	−56	−40	20	4.55
5	Crus I	left	−39	−47	−35	48	4.49
6	VIIb	left	−37	−44	−48	25	3.98

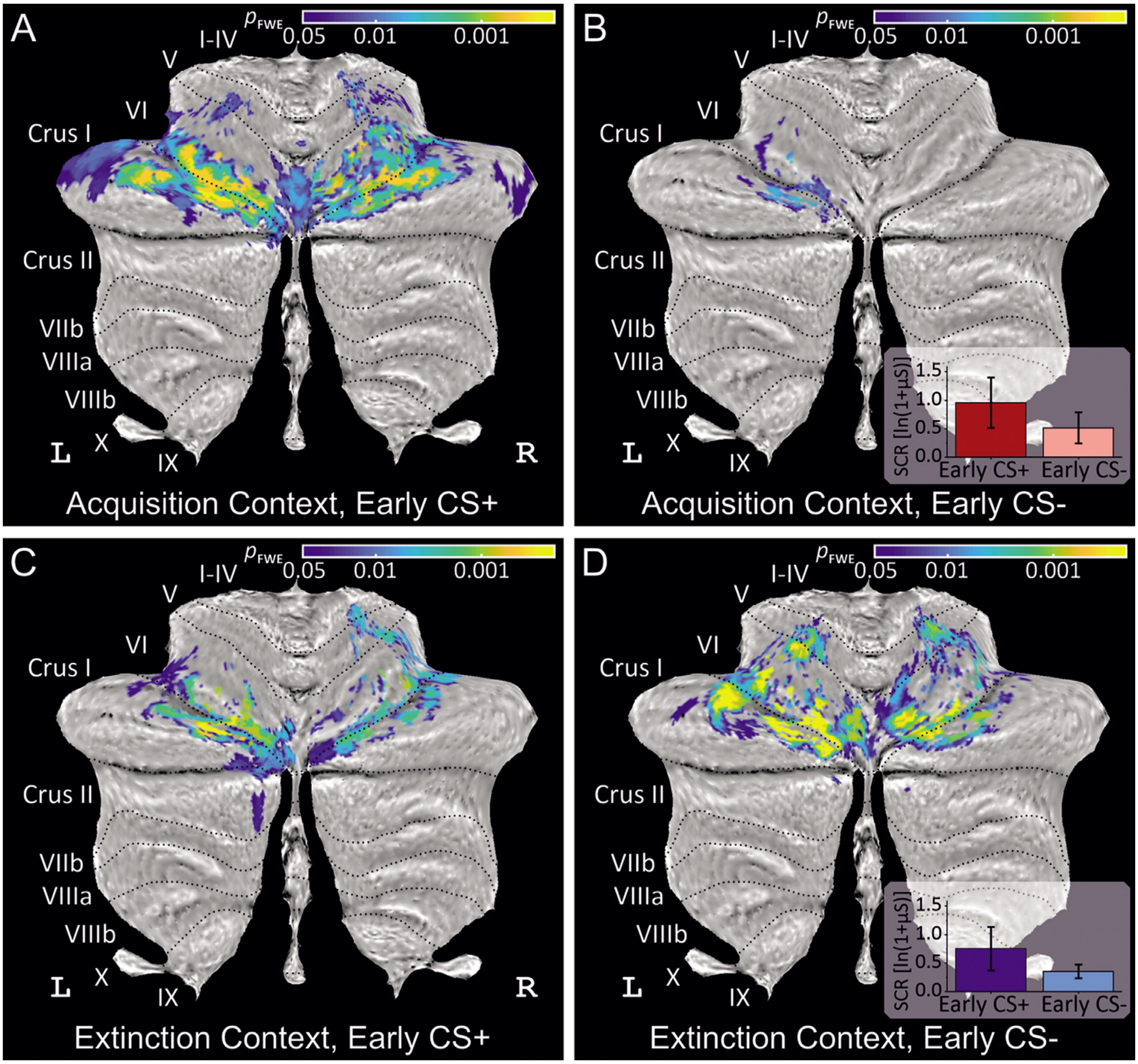

A
B
C
D
p_{FWE} 0.05 0.01 0.001
I-IV
V
VI
Crus I
Crus II
VIIb
VIIIa
VIIIb
X
IX
L
R
Acquisition Context, Early CS+
Acquisition Context, Early CS-
Extinction Context, Early CS+
Extinction Context, Early CS-
SCR [ln(1+µS)]
1.5
1.0
0.5
0.0
Early CS+
Early CS-

Fig. 7. Cerebellar activation related to the CS+ and CS- during early recall. Cerebellar activations during the presentation of stimuli in (**A**) CS+ (acquisition context), (**B**) CS- (acquisition context) in the early recall block, (**C**) CS+ (extinction context), (**D**) CS- (extinction context) in SUIT space projected on a cerebellar flatmap (Diedrichsen and Zotow, 2015). All contrasts are calculated using TFCE and FWE correction (p < 0.05). Insert shows mean skin conductance responses (SCRs) during early. Error bars indicate standard errors (SE). CS = conditioned stimulus; L = left; R = right; SUIT = spatially unbiased atlas template of the cerebellum; TFCE = threshold-free cluster-enhancement; FWE = family-wise error rate.

tion for the no-US in extinction was also significant after application of threshold-free cluster-enhancement (TFCE) at $p < 0.05$ familywise error (FWE) corrected level (Table 8).

During recall we observed a significant effect of the parametric modulation of model-derived prediction values on the CS events in the posterolateral cerebellum at an uncorrected threshold of $p < 0.001$. Effects were observed predominantly in the left lobules Crus I and VI (Fig. 10**A**, **C**; see also Table 7). A small cluster in Crus I remained significant at a threshold of $p < 0.05$ (TFCE-FWE corrected, Table 8). Similar findings were observed performing the same analysis in recall for no-US and model-derived prediction error values (Fig. 10**B**, **D**, Table 7).

4. Discussion

Our main findings were twofold. First, the present fMRI data show that the cerebellum is activated during the context-related recall of previously extinguished learned fear associations. Thus, part of the original associative memory likely remained within the cerebellum following extinction training. Second, as an unexpected incidental finding, we observed that cerebellar activation related to the CS+ was not significantly different from cerebellar activation related to the unreinforced CS- during fear acquisition training. Possible reasons for this lack of differential cerebellar activation will be discussed.

4.1. Lack of differential cerebellar activation comparing CS+ and CS- during fear acquisition training

In the present study, significant cerebellar activations were observed related to both the CS+ and the CS-, which were not significantly different from each other. In differential protocols, learning measures are based on the difference in reactions to the reinforced CS+ and the unreinforced CS- to exclude reactions related to non-associative processes such as orienting responses and habituation (Lonsdorf et al., 2017). The most parsimonious explanation of the present findings is therefore that they reflect a contribution of the cerebellum to these non-associative processes. Cerebellar activations related to the CS+ and the CS- were

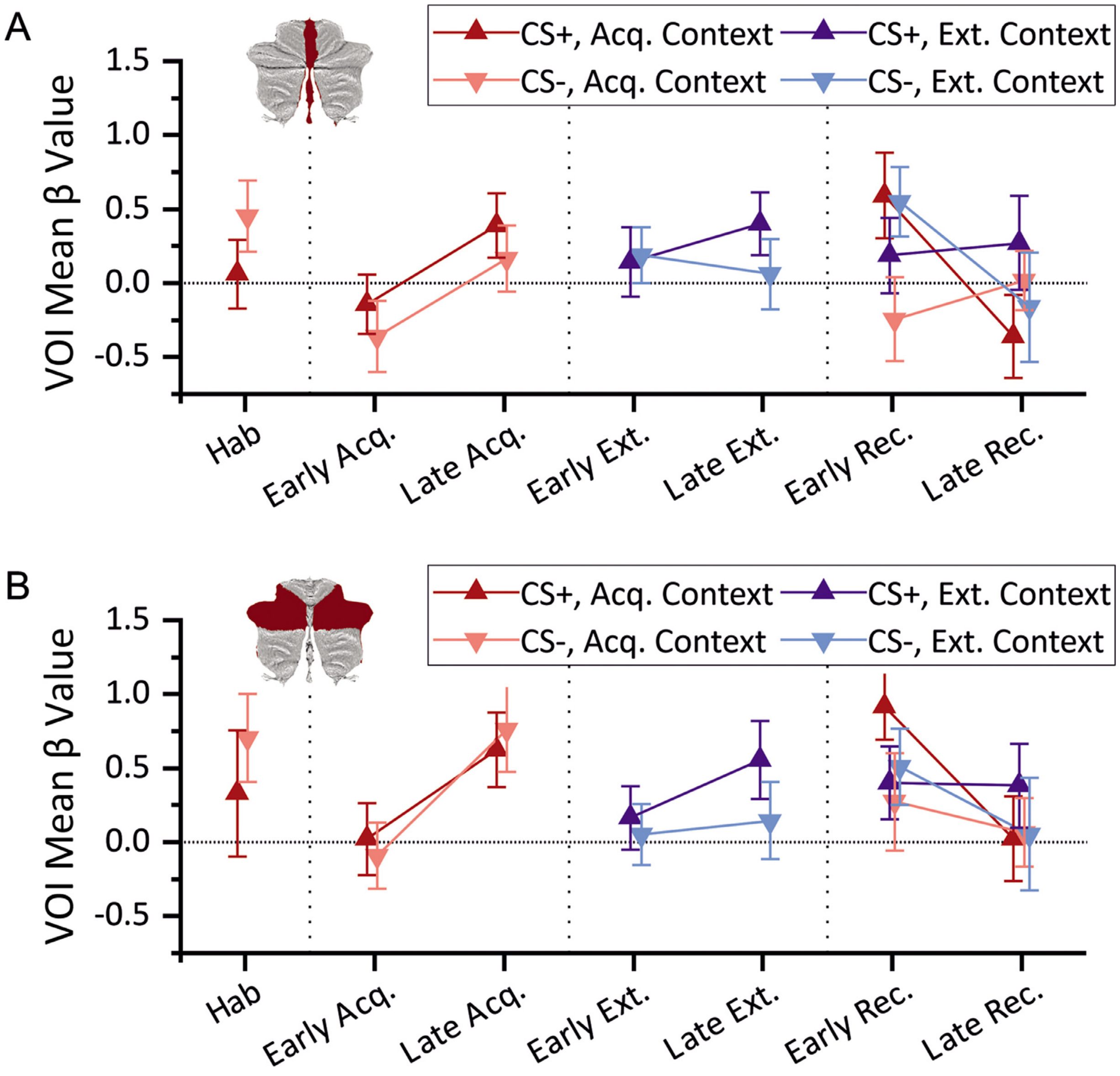


Fig. 8. Mean β values in two VOI. (**A**) vermis and (**B**) lobule VI and Crus I bilaterally on day 1 (habituation, fear acquisition and extinction training) and day 2 (recall) across trials. Dark colors = CS+, light colors = CS-. Trials in acquisition (Acq.) context are shown in red, trials in extinction (Ext.) context are shown in blue. Horizontal lines represent group mean values. Error bars indicate standard error (SE). The two volumes of interest (VOIs) in the vermis and Crus I and lobule VI bilaterally are illustrated in the inserts.

found predominantly in lobules Crus I and VI. These areas overlap with regions in the cerebellum related to emotion processing, but also attention and working memory related processes (Guell et al., 2018; King et al., 2019). Thus, cerebellar activations may be an expression of cerebellar involvement in orienting responses and maintaining attention to the CS.

There is increasing evidence, however, that the CS- is not neutral, and reactions to the CS- go beyond non-associative processes (Lonsdorf et al., 2017). Participants not only learn that the CS+ is followed by the aversive US, but they also learn the non-occurrence of the US following the CS-. It has been proposed that the CS- becomes a learned safety cue predicting the absence of a harmful event (Christianson et al., 2012; Grasser and Jovanovic, 2021; Labrenz et al., 2015). Thus, the cerebellum may also be involved in associative learning-processes related to the CS-, for example as a safety cue. The absence of an aversive stimulus is rewarding (Kalisch et al., 2019) and it has been shown that the cerebellum receives reward signals (Carta et al., 2019; Wagner et al., 2017; Wagner and Luo, 2020). As yet, however, it is still debated whether safety learning occurs as a form of reward learning. Future studies are needed to test the hypotheses that cerebellar activation related to the CS- go beyond non-associative processes, and include learned associations related to the CS-.

Most previous studies in the literature, however, observed differential activation in the cerebellar cortex with significantly stronger activations related to the CS+ compared to the CS- (Ernst et al., 2019; Lange et al., 2015). Differences in parameters of the applied conditioning paradigms may explain the lack of differential activation in the present study. First, the US was applied to the lower leg and not to the hand. Harmful stimuli applied to the hand are more salient compared to the face (Schmidt et al., 2020). Likewise, harmful stimuli applied to the hand are likely more salient compared to the lower limb. Increased US salience leads to stronger associative learning (Lonsdorf et al., 2017) and may result in stronger cerebellar activations related to the CS+.

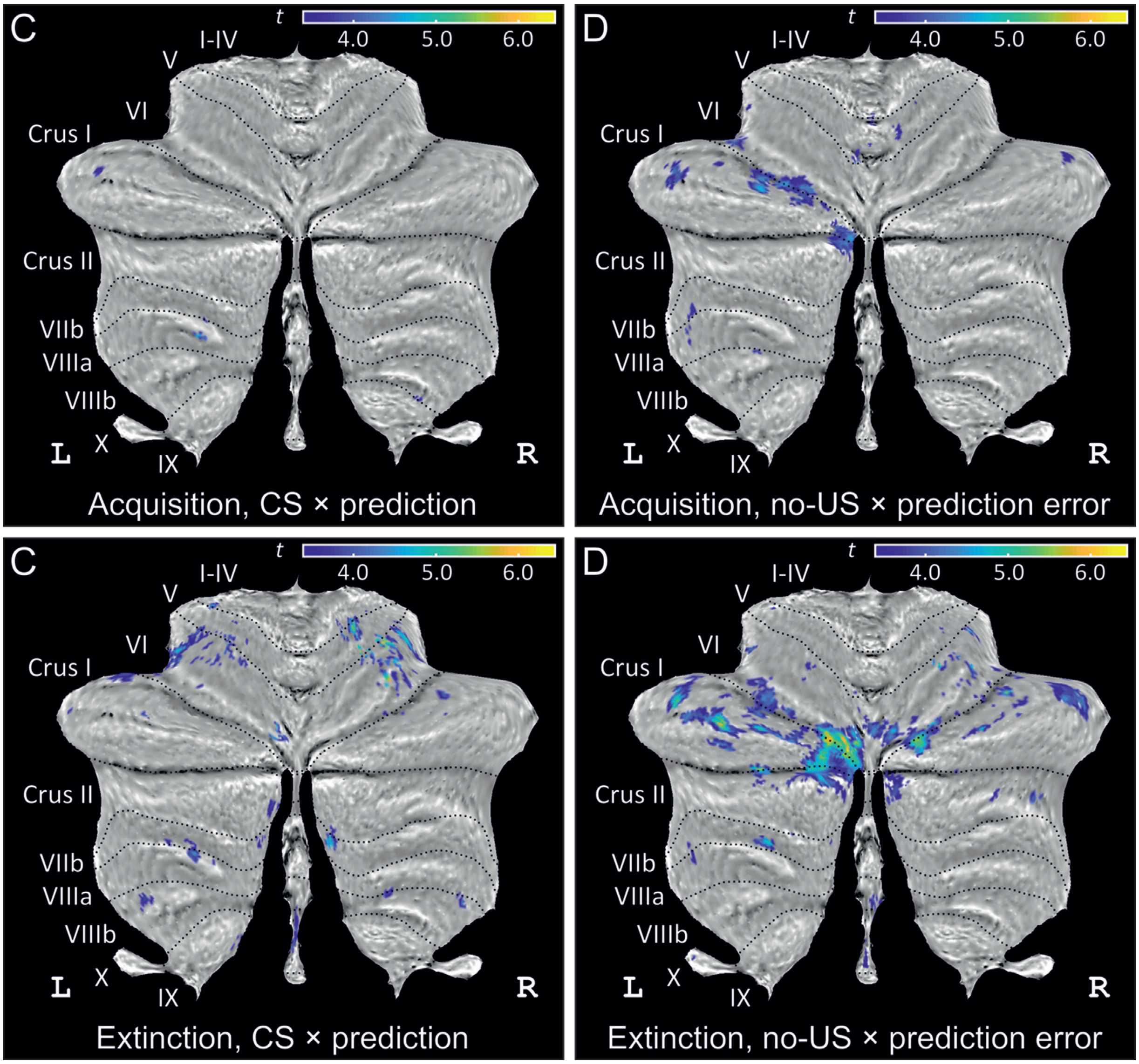

C
t
4.0
5.0
6.0
I-IV
V
VI
Crus I
Crus II
VIIb
VIIIa
VIIIb
X
IX
L
R
Acquisition, CS × prediction
D
Acquisition, no-US × prediction error
C
Extinction, CS × prediction
D
Extinction, no-US × prediction error

Fig. 9. Group main effects of parametric modulation with learning model-derived data during fear acquisition and extinction training. (**A, C**) Parametric modulation of CS events with individual mean prediction values, and (**B, D**) parametric modulation of omission of US events at CS termination (no-US) with individual absolute mean prediction error values. Activation maps are displayed at a trend level of $p < 0.001$, uncorrected, on a cerebellar flatmap.

Furthermore, in the present study, cues were shown in a complex context. It is conceivable that participants learned not only the association between the cue/context and the US, but also between the cue and the context. The cerebellum is known to contribute to learning of cognitive stimulus-stimulus associations (Drepper et al., 1999). The presentation of the context had an onset and an end. Participants may have learned that the presentation of the context is followed by turning on a light, independent of whether this is followed by an electric shock or not. Learning this context-cue cognitive association may have resulted in cerebellar activations related to both the CS+ and the CS-, which may have hampered the detection of differential activations related to the CS-US association. Despite this lack of differential activation, significant effects of parametric modulation with model-derived prediction error values were found in the cerebellum in the extinction training and on a trend level in the acquisition training. Prediction errors drive associative learning (Holland and Schiffino, 2016; Rescorla and Wagner, 1972), and parametric modulation effects suggest that fear and extinction learning related processes have taken place in the cerebellum. These findings agree with a previous study of our group that found that the cerebellum contributes to the processing of predictions and prediction errors in fear conditioning (Ernst et al., 2019). Regardless of possible reasons of the lack of significant differential activations, the present study shows that it is worthwhile to look at the individual contrasts related to the reinforced and unreinforced cues, in addition to the differential contrasts. Although part of the activations related to unreinforced cues are likely due to non-associative processes, cerebellar activations related to learning of a safety signal or other learned associations related to the CS- may also play a role.

4.2. Cerebellum contributes to context-related processes of extinction

The main aim of the present study was to show that associative fear memory in the cerebellum is not fully erased during extinction training. We believe that our findings agree with this assumption. During recall, strongest cerebellar activation was seen during early recall and presentation of the CS+ in the acquisition context. Cerebellar activation was accompanied by significantly increased SCRs as an indication of return of fear, i.e., renewal. Context-related return of fear during recall was further supported by modeling data. During recall parametric modulation of the CS with the US prediction values calculated by our DNN

Table 8
Parametric modulation of learning model-derived prediction parameters after TFCE and FWE correction. Displayed are all clusters that survive TFCE and FWE correction thresholded at $p < 0.05$ without cluster size limit. In each cluster, up to three maxima are listed separated by ≥ 8 mm.

Index	Location (lobule)	Side	SUIT coordinates/mm			Cluster size/mm^3	p_{FWE}	*TFCE*
CS × prediction in fear acquisition training								
	no significant clusters							
CS × prediction in extinction training								
	no significant clusters							
CS × prediction in recall								
1	Crus I	left	−35	−47	−35	4	0.044	750.2
no-US × prediction error in fear acquisition training								
	no significant clusters							
no-US × prediction error in extinction training								
1	Extended cluster	left Crus I (4808), right VI (1837), left VI (1807), right Crus I (1602), vermal VI (629), left Crus II (555), right Crus II (204), left V (62), vermal Crus II (58), vermal Crus I (17), gray matter (5), right V (3)						
	VI	left	−18	−78	−20	11,587	<0.001	2216.6
	VI	left	−7	−81	−21		<0.001	2160.3
	Crus I	left	−15	−81	−27		<0.001	2086.1
2	Crus I	left	−27	−67	−38	109	0.011	983.4
3	Extended cluster	left VI (137)						
	VI	left	−26	−59	−17	137	0.032	812.4
	VI	left	−17	−63	−17		0.041	779.7
4	VI	right	9	−69	−14	31	0.045	754.9
5	Crus I	right	25	−85	−34	27	0.046	750.3
6	Crus II	left	−30	−78	−40	3	0.049	743.1
7	Crus I	left	−43	−77	−41	2	0.050	740.8
no-US × prediction error in recall								
	no significant clusters							

reinforcement learning model showed significantly modulated activation of the posterolateral cerebellum (lobules Crus I and VI). Since the prediction values differed based on context during recall, data suggest that the cerebellum is involved in context-related recall of learned fear associations. Findings agree with observations in patients with cerebellar disease which affected parts of the posterolateral hemisphere. These patients showed a lack of renewal in an eyeblink conditioning paradigm (Steiner et al., 2019).

A prerequisite for context-related return of previously extinguished fear responses is that part of the original fear memory is retained (Bouton and King, 1983). Cerebellar activation in the cerebellar cortex during recall suggests that memory is retained in the cerebellar cortex. This is at variance with a recording study in fish showing that acquisition related plasticity at the Purkinje cell is reversed during extinction (Yoshida and Kondo, 2012). These findings agree with recording studies in the eyeblink conditioning paradigms in rodents (Jirenhed et al., 2007). In eyeblink conditioning, however, plastic changes during acquisition training are not limited to the parallel fiber/Purkinje cell synapse, but also occur at granule cells and interneurons (De Zeeuw et al., 2021; Gao et al., 2012). This may equally be the case in fear conditioning. Granule cells and interneurons may be places where memory is retained. In the eyeblink conditioning literature, it has also been proposed that memory is retained in the cerebellar nuclei allowing for fast relearning in the cerebellar cortex during reacquisition (and therefore explaining saving effects) (Medina et al., 2002). The remaining memory may be under the inhibitory control of the cerebral fear extinction network mediated via the amygdala and pons as outlined in more detail in the introduction (Hu et al., 2015; Robleto and Thompson, 2008). The cerebellum on the other hand has known anatomical connections with the vmPFC and the hippocampus and may also modulate context-related processes in extinction learning (Bostan et al., 2018).

Cerebellar activation in recall, however, was also present in the extinction context, and also related to the CS-. Fear acquisition and extinction training were performed on one day and recall on the subsequent day. A change in context can also be a change in time. Thus, cerebellar activation may also be explained by an ABC renewal effect (fear acquisition training in context A, extinction training in context B, recall in context C) (Hermann et al., 2016). Activations related to the CS- suggest that associations related to the CS- (e.g., learned safety signals) are also context-dependent, but this needs to be confirmed and closer investigated in future studies. Finally, higher cerebellar activation related to the CS- in the extinction compared to the acquisition context may reflect the fact that participants expected a change of contingencies.

4.3. Limitations

SCRs are known to show intra- and interindividual variability (Dawson et al., 2007; Fowles and Rosenberry, 1973; Maulsby and Edelberg, 1960). This was also the case in the present study. Therefore, modeling data were based on group data rather than individual data. Inclusion of additional outcome parameters, e.g., changes in pupil size (Leuchs et al., 2017, 2019) in future studies may allow for more robust trial-by-trial individual behavioral data and improve modeling of predictive trial-by-trial values in individual participants.

Cerebellar activations in fear conditioning studies were most robust in the posterolateral cerebellum, which is in very good agreement with the previous literature. Activations of the lateral cerebellum may be most strongly related to more cognitive aspects of the learned associations, which have not been assessed on a trial-by-trial basis. Furthermore, as discussed above, cerebellar activations may not only be related to the CS+/US associations.

Furthermore, although cerebellar activation related to the CS+ was significantly higher compared to CS- in the acquisition but not the extinction context during early recall, no significant cerebellar activation was observed comparing the CS+ in the acquisition and extinction context. Results need to be confirmed in a larger group of participants using an optimized paradigm, e.g., by application of the US to the hand and by continuous presentations of the contexts. Possible reasons for lack of statistical difference comparing cerebellar activations related to the CS+ and CS- in acquisition learning have been outlined above. Lack of power may be an additional reason why significant differential activations were not observed.

Finally, the design worked in general, but not in each phase and for each measure. An additional look at other prominent fear related brain structures would likely add further signs of successful learning.

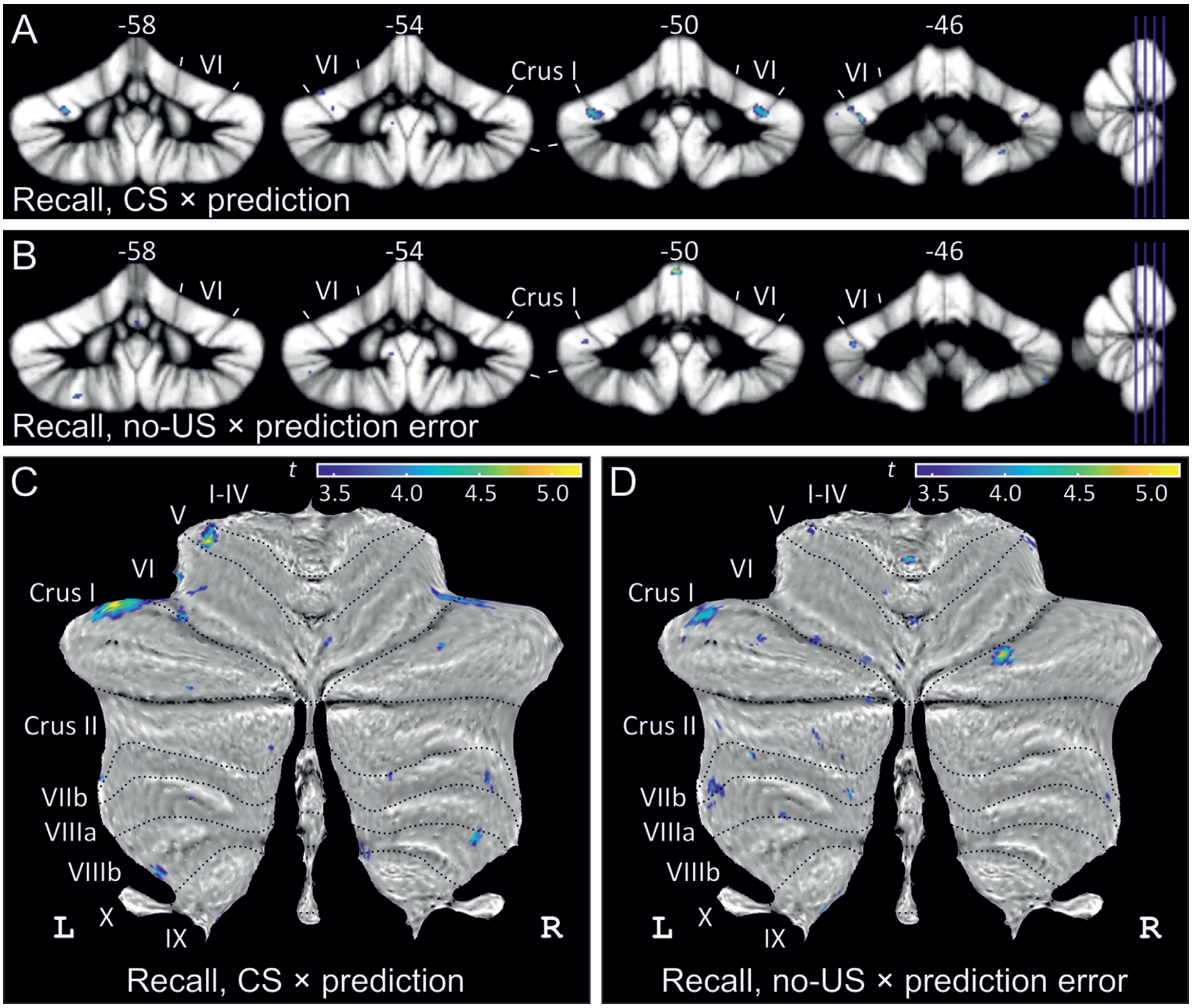


Fig. 10. Group main effects of parametric modulation with learning model-derived prediction parameters during recall. (**A, C**) Parametric modulation of CS events with individual mean prediction values, and (**B, D**) parametric modulation of omission of US events at CS termination (no-US) with individual mean absolute prediction error values. Activation maps are displayed at a trend level of $p < 0.001$, uncorrected, on selected coronal slices of the cerebellum (**A, B**), and on a cerebellar flatmap (**C, D**).

5. Conclusions

Cerebellar activation was present during context-related recall of previously extinguished fear associations, which agrees with the assumption that part of the original associative memory is retained in the cerebellum. As an unexpected side-effect we found lack of differential cerebellar activations related to the CS+ and CS-. Cerebellar activation related to the CS- may be related to non-associative processes, such as orienting responses, or learning of CS- related associations, e.g., in the context of safety cues.

Funding

This work was supported by a grant from the German Research Foundation (DFG; project number 316,803,389 – SFB 1280) to D.T. and H.H.Q. (subproject A05), C.J.M. (subproject A09) and S.C. (subproject F01). The 7-Tesla MRI system at the Erwin L. Hahn Institute was funded by the German Research Foundation (DFG; project number 432,657,511).

Data and code availability statement

All MATLAB and Python source code used in this paper are available upon direct request to the corresponding author. The consent form that participants signed does not allow us to share the raw data publicly, but it can be made available upon request to interested researchers through a data sharing agreement.

Credit authorship contribution statement

Giorgi Batsikadze: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Nicolas Diekmann:** Methodology, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. **Thomas Michael Ernst:** Conceptualization, Validation, Data curation, Supervision, Methodology, Formal analysis, Investigation, Resources, Visualization, Software, Writing – original draft, Writing – review & editing. **Michael Klein:** Data curation, Investigation. **Stefan Maderwald:** Resources, Supervision. **Cornelius Deuschl:** Resources. **Christian Josef Merz:** Conceptualization, Methodology, Writing – review & editing. **Sen Cheng:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – review & editing. **Harald H. Quick:** Resources, Funding acquisition. **Dagmar Timmann:** Conceptualization, Funding acquisition, Project administration, Supervision, Methodology, Writing – review & editing.

Acknowledgments

The authors like to thank P.J. Koopmans and V. Pfaffenroth for the advice and help reconstruction the 3D-EPI, P. Ehses for contributing the code for the 3D-EPI reconstruction, T. Otto for his work on SCR analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119080.

References

Llerena, A., de la Rubia, A., Penas-Lledo, E.M., Diaz, F.J., de Leon, J., 2002. Schizophrenia and tobacco smoking in a Spanish psychiatric hospital. Schizophr. Res. 58, 323–327.
Akritas, M.G., 1990. The rank transform method in some two-factor designs. J. Am. Stat. Assoc. 85, 73–78.
Apps, R., Strata, P., 2015. Neuronal circuits for fear and anxiety - the missing link. Nat. Rev. Neurosci. 16, 642.
Barad, M., 2006. Is extinction of fear erasure or inhibition? Why both, of course. Learn. Mem. 13, 108–109.
Bostan, A.C., Dum, R.P., Strick, P.L., 2018. Functional anatomy of basal ganglia circuits with the cerebral cortex and the cerebellum. Prog. Neurol. Surg. 33, 50–61.
Boucsein, W., 2012. Electrodermal Activity, 2nd ed. Springer, New York.
Bouton, M.E., 2004. Context and behavioral processes in extinction. Learn. Mem. 11, 485–494.
Bouton, M.E., King, D.A., 1983. Contextual control of the extinction of conditioned fear: tests for the associative value of the context. J. Exp. Psychol. Anim. Behav. Process. 9, 248–265.
Breuer, F.A., Blaimer, M., Heidemann, R.M., Mueller, M.F., Griswold, M.A., Jakob, P.M., 2005. Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. Magn. Reson. Med. 53, 684–691.
Brunner, E., Dette, H., Munk, A., 1997. Box-type approximations in nonparametric factorial designs. J. Am. Stat. Assoc. 92, 1494–1502.
Brunner, E., Domhof, S., Langer, F., 2002. Nonparametric Analysis of Longitudinal Data in Factorial Experiments. J. Wiley, New York, NY.
Brunner, E., Munzel, U., Puri, M.L., 1999. Rank-score tests in factorial designs with repeated measures. J. Multivar. Anal. 70, 286–317.
Carta, I., Chen, C.H., Schott, A.L., Dorizan, S., Khodakhah, K., 2019. Cerebellar modulation of the reward circuitry and social behavior. Science 363 eaav0581.
Christianson, J.P., Fernando, A.B., Kazama, A.M., Jovanovic, T., Ostroff, L.E., Sangha, S., 2012. Inhibition of fear by learned safety signals: a mini-symposium review. J. Neurosci. 32, 14118–14124.
Craske, M.G., Hermans, D., Vervliet, B., 2018. State-of-the-art and future directions for extinction as a translational model for fear and anxiety. Philos. Trans. R. Soc. B Biol. Sci. 373.
Dawson, M.E., Schell, A.M., Filion, D.L., Berntson, G.G., Cacioppo, J.T., Tassinary, L.G., Berntson, G., 2007. The Electrodermal System. Handbook of Psychophysiology. Cambridge University Press, New York, NY, US, pp. 157–181.
De Zeeuw, C.I., Lisberger, S.G., Raymond, J.L., 2021. Diversity and dynamism in the cerebellum. Nat. Neurosci. 24, 160–167.
Diedrichsen, J., 2006. A spatially unbiased atlas template of the human cerebellum. Neuroimage 33, 127–138.
Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M., Ivry, R.B., 2019. Universal transform or multiple functionality? Understanding the contribution of the human cerebellum across task domains. Neuron 102, 918–928.
Diedrichsen, J., Zotow, E., 2015. Surface-based display of volume-averaged cerebellar imaging data. PLoS One 10, e0133402.
Dietrichs, E., Haines, D.E., 1989. Interconnections between hypothalamus and cerebellum. Anat. Embryol. 179, 207–220 Berl..
Drepper, J., Timmann, D., Kolb, F.P., Diener, H.C., 1999. Non-motor associative learning in patients with isolated degenerative cerebellar disease. Brain 122 (Pt 1), 87–97.
Dubois, C.J., Fawcett-Patel, J., Katzman, P.A., Liu, S.J., 2020. Inhibitory neurotransmission drives endocannabinoid degradation to promote memory consolidation. Nat. Commun. 11, 6407.
Ernst, T.M., Brol, A.E., Gratz, M., Ritter, C., Bingel, U., Schlamann, M., Maderwald, S., Quick, H.H., Merz, C.J., Timmann, D., 2019. The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. Elife 8, e46831.
Ernst, T.M., Thurling, M., Muller, S., Kahl, F., Maderwald, S., Schlamann, M., Boele, H.J., Koekkoek, S.K.E., Diedrichsen, J., De Zeeuw, C.I., Ladd, M.E., Timmann, D., 2017. Modulation of 7 T fMRI Signal in the cerebellar cortex and nuclei during acquisition, extinction, and reacquisition of conditioned eyeblink responses. Hum. Brain Mapp. 38, 3957–3974.
Farley, S.J., Radley, J.J., Freeman, J.H., 2016. Amygdala modulation of cerebellar learning. J. Neurosci. 36, 2190–2201.
Fischer, H., Andersson, J.L., Furmark, T., Fredrikson, M., 2000. Fear conditioning and brain activity: a positron emission tomography study in humans. Behav. Neurosci. 114, 671–680.
Fowles, D.C., Rosenberry, R., 1973. Effects of epidermal hydration on skin potential responses and levels. Psychophysiology 10, 601–611.
Frontera, J.L., Baba Aissa, H., Sala, R.W., Mailhes-Hamon, C., Georgescu, I.A., Léna, C., Popa, D., 2020. Bidirectional control of fear memories by cerebellar neurons projecting to the ventrolateral periaqueductal grey. Nat. Commun. 11, 5207.
Gao, Z., van Beugen, B.J., De Zeeuw, C.I., 2012. Distributed synergistic plasticity and cerebellar learning. Nat. Rev. Neurosci. 13, 619–635.
Graham, B.M., Milad, M.R., 2013. Blockade of estrogen by hormonal contraceptives impairs fear extinction in female rats and women. Biol. Psychiatry 73, 371–378.
Grasser, L.R., Jovanovic, T., 2021. Safety learning during development: implications for development of psychopathology. Behav. Brain Res. 408, 113297.
Guell, X., Gabrieli, J.D.E., Schmahmann, J.D., 2018. Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort. Neuroimage 172, 437–449.
Guell, X., Schmahmann, J., 2020. Cerebellar functional anatomy: a didactic summary based on human fMRI evidence. Cerebellum 19, 1–5.
Han, J.K., Kwon, S.H., Kim, Y.G., Choi, J., Kim, J.I., Lee, Y.S., Ye, S.K., Kim, S.J., 2021. Ablation of STAT3 in Purkinje cells reorganizes cerebellar synaptic plasticity in long-term fear memory network. Elife 10, e63291.
Heath, R.G., Harper, J.W., 1974. Ascending projections of the cerebellar fastigial nucleus to the hippocampus, amygdala, and other temporal lobe sites: evoked potential and histological studies in monkeys and cats. Exp. Neurol. 45, 268–287.
Hermann, A., Stark, R., Milad, M.R., Merz, C.J., 2016. Renewal of conditioned fear in a novel context is associated with hippocampal activation and connectivity. Soc. Cogn. Affect. Neurosci. 11, 1411–1421.
Herry, C., Ferraguti, F., Singewald, N., Letzkus, J.J., Ehrlich, I., Luthi, A., 2010. Neuronal circuits of fear extinction. Eur. J. Neurosci. 31, 599–612.
Hesslow, G., Ivarsson, M., 1996. Inhibition of the inferior olive during conditioned responses in the decerebrate ferret. Exp. Brain Res. 110, 36–46.
Holland, P.C., Schiffino, F.L., 2016. Mini-review: prediction errors, attention and associative learning. Neurobiol. Learn. Mem. 131, 207–215.
Hu, C., Zhang, L.B., Chen, H., Xiong, Y., Hu, B., 2015. Neurosubstrates and mechanisms underlying the extinction of associative motor memory. Neurobiol. Learn. Mem. 126, 78–86.
Hull, C., 2020. Prediction signals in the cerebellum: beyond supervised motor learning. Elife 9, e54073.
Inoue, L., Ernst, T.M., Ferber, I.I., Merz, C.J., Timmann, D., Batsikadze, G., 2020. Interaction of fear conditioning with eyeblink conditioning supports the sensory gating hypothesis of the amygdala in men. eNeuro 7 ENEURO.0128-0120.2020.
Jirenhed, D.A., Bengtsson, F., Hesslow, G., 2007. Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace. J. Neurosci. 27, 2493–2502.
Kalisch, R., Gerlicher, A.M.V., Duvarci, S., 2019. A dopaminergic basis for fear extinction. Trends Cogn. Sci. 23, 274–277.
Kattoor, J., Thurling, M., Gizewski, E.R., Forsting, M., Timmann, D., Elsenbruch, S., 2014. Cerebellar contributions to different phases of visceral aversive extinction learning. Cerebellum 13, 1–8.
Kim, J.J., Jung, M.W., 2006. Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. Neurosci. Biobehav. Rev. 30, 188–202.
Kim, O.A., Ohmae, S., Medina, J.F., 2020. A cerebello-olivary signal for negative prediction error is sufficient to cause extinction of associative motor learning. Nat. Neurosci. 23, 1550–1554.
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. Nat. Neurosci. 22, 1371–1378.
Kinner, V.L., Merz, C.J., Lissek, S., Wolf, O.T., 2016. Cortisol disrupts the neural correlates of extinction recall. Neuroimage 133, 233–243.
Koutsikou, S., Crook, J.J., Earl, E.V., Leith, J.L., Watson, T.C., Lumb, B.M., Apps, R., 2014. Neural substrates underlying fear-evoked freezing: the periaqueductal grey-cerebellar link. J. Physiol. 592, 2197–2213.
Labrenz, F., Icenhour, A., Thurling, M., Schlamann, M., Forsting, M., Timmann, D., Elsenbruch, S., 2015. Sex differences in cerebellar mechanisms involved in pain-related safety learning. Neurobiol. Learn. Mem. 123, 92–99.
Lange, I., Kasanova, Z., Goossens, L., Leibold, N., De Zeeuw, C.I., van Amelsvoort, T., Schruers, K., 2015. The anatomy of fear learning in the cerebellum: a systematic meta–analysis. Neurosci. Biobehav. Rev. 59, 83–91.
Larrauri, J.A., Schmajuk, N.A., 2008. Attentional, associative, and configural mechanisms in extinction. Psychol. Rev. 115, 640–676.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Leuchs, L., Schneider, M., Czisch, M., Spoormaker, V.I., 2017. Neural correlates of pupil dilation during human fear learning. Neuroimage 147, 186–197.
Leuchs, L., Schneider, M., Spoormaker, V.I., 2019. Measuring the conditioned response: a comparison of pupillometry, skin conductance, and startle electromyography. Psychophysiology 56, e13283.
Lin, L.J., 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. Mach. Learn. 8, 293–321.
Liu, W., Zhang, Y., Yuan, W., Wang, J., Li, S., 2012. A direct hippocampo-cerebellar projection in chicken. Anat. Rec. 295, 1311–1320 Hoboken.
Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Romer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas, J.M.P., Merz, C.J., 2017. Don't fear 'fear conditioning': methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci. Biobehav. Rev. 77, 247–285.
Lovibond, S.H., Lovibond, P.F.Psychology Foundation of, A., 1995. Manual for the Depression Anxiety Stress Scales.
Manto, M., Bower, J.M., Conforto, A.B., Delgado-Garcia, J.M., da Guarda, S.N., Gerwig, M., Habas, C., Hagura, N., Ivry, R.B., Marien, P., Molinari, M., Naito, E., Nowak, D.A., Oulad Ben Taib, N., Pelisson, D., Tesche, C.D., Tilikete, C., Timmann, D., 2012. Consensus paper: roles of the cerebellum in motor control–the diversity of ideas on cerebellar involvement in movement. Cerebellum 11, 457–487.

Maren, S., Phan, K.L., Liberzon, I., 2013. The contextual brain: implications for fear conditioning, extinction and psychopathology. Nat. Rev. Neurosci. 14, 417–428.

Maschke, M., Schugens, M., Kindsvater, K., Drepper, J., Kolb, F.P., Diener, H.C., Daum, I., Timmann, D., 2002. Fear conditioned changes of heart rate in patients with medial cerebellar lesions. J. Neurol. Neurosurg. Psychiatry 72, 116–118.

Maulsby, R.L., Edelberg, R., 1960. The interrelationship between the galvanic skin response, basal resistance, and temperature. J. Comp. Physiol. Psychol. 53, 475–479.

Medina, J.F., Nores, W.L., Mauk, M.D., 2002. Inhibition of climbing fibres is a signal for the extinction of conditioned eyelid responses. Nature 416, 330–333.

Merz, C.J., Kinner, V.L., Wolf, O.T., 2018. Let's talk about sex … differences in human fear conditioning. Curr. Opin. Behav. Sci. 23, 7–12.

Middleton, F.A., Strick, P.L., 2001. Cerebellar projections to the prefrontal cortex of the primate. J. Neurosci. 21, 700–712.

Milad, M.R., Quirk, G.J., 2012. Fear extinction as a model for translational neuroscience: ten years of progress. Annu. Rev. Psychol. 63, 129–151.

Milad, M.R., Wright, C.I., Orr, S.P., Pitman, R.K., Quirk, G.J., Rauch, S.L., 2007. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biol. Psychiatry 62, 446–454.

Molchan, S.E., Sunderland, T., McIntosh, A.R., Herscovitch, P., Schreurs, B.G., 1994. A functional anatomical study of associative learning in humans. Proc. Natl. Acad. Sci. USA 91, 8122–8126.

Myers, K.M., Davis, M., 2007. Mechanisms of fear extinction. Mol. Psychiatry 12, 120–150.

Newman, P.P., Reza, H., 1979. Functional relationships between the hippocampus and the cerebellum: an electrophysiological study of the cat. J. Physiol. 287, 405–426.

Noguchi, K., Gel, Y.R., Brunner, E., Konietschke, F., 2012. nparLD: anRSoftware package for the nonparametric analysis of longitudinal data in factorial experiments. J. Stat. Softw. 50, 1–23.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Onat, F., Cavdar, S., 2003. Cerebellar connections: hypothalamus. Cerebellum 2, 263–269.

Pavlov, I.P., 1927. Conditioned reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex. Oxford Univ. Press, Oxford, England.

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, 897–905.

Pineles, S.L., Orr, M.R., Orr, S.P., 2009. An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. Psychophysiology 46, 984–995.

Ploghaus, A., Tracey, I., Gati, J.S., Clare, S., Menon, R.S., Matthews, P.M., Rawlins, J.N., 1999. Dissociating pain from its anticipation in the human brain. Science 284, 1979–1981.

Popa, L.S., Ebner, T.J., 2019. Cerebellum, Predictions and Errors. Front. Cell. Neurosci. 12.

Rescorla, R.A., Wagner, A., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), *Classical conditioning II: Current research and theory*. Appleton-Century-Crofts, New York, pp. 64–99.

Robleto, K., Poulos, A.M., Thompson, R.F., 2004. Brain mechanisms of extinction of the classically conditioned eyeblink response. Learn. Mem. 11, 517–524.

Robleto, K., Thompson, R.F., 2008. Extinction of a classically conditioned response: red nucleus and interpositus. J. Neurosci. 28, 2651–2658.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back--propagating errors. Nature 323, 533–536.

Sacchetti, B., Baldi, E., Lorenzini, C.A., Bucherelli, C., 2002. Cerebellar role in fear-conditioning consolidation. Proc. Natl. Acad. Sci. USA 99, 8406–8411.

Sacchetti, B., Scelfo, B., Tempia, F., Strata, P., 2004. Long-term synaptic changes induced in the cerebellar cortex by fear conditioning. Neuron 42, 973–982.

T. Schaul, J. Quan, I. Antonoglou, D. Silver, 2015. Prioritized experience replay. arXiv preprint arXiv:1511.05952.

Schmahmann, J.D., 2019. The cerebellum and cognition. Neurosci. Lett. 688, 62–75.

Schmidt, K., Forkmann, K., Elsenbruch, S., Bingel, U., 2020. Enhanced pain-related conditioning for face compared to hand pain. PLoS One 15, e0234160.

Shah, D.A., Madden, L.V., 2004. Nonparametric analysis of ordinal data in designed factorial experiments. Phytopathology 94, 33–43.

Sokolov, A.A., Miall, R.C., Ivry, R.B., 2017. The cerebellum: adaptive prediction for movement and cognition. Trends Cogn. Sci. 21, 313–332 Regul. Ed..

Steiner, K.M., Gisbertz, Y., Chang, D.I., Koch, B., Uslar, E., Claassen, J., Wondzinski, E., Ernst, T.M., Goricke, S.L., Siebler, M., Timmann, D., 2019. Extinction and renewal of conditioned eyeblink responses in focal cerebellar disease. Cerebellum 18, 166–177.

Supple, W.F., Leaton, R.N., 1990a. Cerebellar vermis: essential for classically conditioned bradycardia in the rat. Brain Res. 509, 17–23.

Supple, W.F., Leaton, R.N., 1990b. Lesions of the cerebellar vermis and cerebellar hemispheres: effects on heart rate conditioning in rats. Behav. Neurosci. 104, 934–947.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction, 2nd MIT Press.

Teeuwisse, W.M., Brink, W.M., Webb, A.G., 2012. Quantitative assessment of the effects of high-permittivity pads in 7 Tesla MRI of the brain. Magn. Reson. Med. 67, 1285–1293.

Thürling, M., Kahl, F., Maderwald, S., Stefanescu, R.M., Schlamann, M., Boele, H.J., De Zeeuw, C.I., Diedrichsen, J., Ladd, M.E., Koekkoek, S.K., Timmann, D., 2015. Cerebellar cortex and cerebellar nuclei are concomitantly activated during eyeblink conditioning: a 7T fMRI study in humans. J. Neurosci. 35, 1228–1239.

Utz, A., Thurling, M., Ernst, T.M., Hermann, A., Stark, R., Wolf, O.T., Timmann, D., Merz, C.J., 2015. Cerebellar vermis contributes to the extinction of conditioned fear. Neurosci. Lett. 604, 173–177.

Venables, P.H., Christie, M.J., 1980. Electrodermal activity. In: Martin, I., Venables, P.H. (Eds.), Techniques in Psychophysiology. Wiley, New York, pp. 3–67.

Wagner, M.J., Kim, T.H., Savall, J., Schnitzer, M.J., Luo, L., 2017. Cerebellar granule cells encode the expectation of reward. Nature 544, 96–100.

Wagner, M.J., Luo, L., 2020. Neocortex-cerebellum circuits for cognitive processing. Trends Neurosci. 43, 42–54.

Walther, T., Diekmann, N., Vijayabaskaran, S., Donoso, J.R., Manahan-Vaughan, D., Wiskott, L., Cheng, S., 2021. Context-dependent extinction learning emerging from raw sensory inputs: a reinforcement learning approach. Sci. Rep. 11, 2713.

Watson, T.C., Becker, N., Apps, R., Jones, M.W., 2014. Back to front: cerebellar connections and interactions with the prefrontal cortex. Front. Syst. Neurosci. 8, 4.

Yoshida, M., Kondo, H., 2012. Fear conditioning-related changes in cerebellar Purkinje cell activities in goldfish. Behav. Brain Funct. 8, 52.