Reviewed Preprint
v1 • January 31, 2025
Not revised

Reviewed Preprint
v2 • April 14, 2026
Revised by authors

For correspondence:
enzo.nio@uk-essen.de

Competing interests: No competing interests declared

Funding: See page 72

Reviewing editor: Andre F Marquand, Radboud University Nijmegen, Netherlands

Human cerebellum and ventral tegmental area interact during extinction of learned fear

Enzo Nio[1,2] ✉, Patrick Pais Pereira[1,2], Nicolas Diekmann[3], Mykola Petrenko[1,2], Alice Doubliez[1,2], Thomas M Ernst[1,2], Giorgi Batsikadze[1,2], Stefan Maderwald[2], Cornelius Deuschl[4,5], Metin Üngör[6], Sen Cheng[3], Christian J Merz[7], Harald H Quick[2,8], Dagmar Timmann[1,2]

[1]Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), Essen University Hospital, University of Duisburg-Essen, Essen, Germany • [2]Erwin L. Hahn Institute for Magnetic Resonance Imaging, University of Duisburg-Essen, Essen, Germany; • [3]Institute for Neural Computation, Faculty of Computer Science, Ruhr University Bochum, Bochum, Germany • [4]Institute for Diagnostic and Interventional Radiology and Neuroradiology, Essen University Hospital, University of Duisburg-Essen, Essen, Germany • [5]Institute for Diagnostic and Interventional Radiology, Neuroradiology and Nuclear Medicine, University Hospital Knappschaftskrankenhaus Bochum, Ruhr University Bochum, Bochum, Germany • [6]Department of Psychology, Philipps-University Marburg, Marburg, Germany • [7]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Ruhr University Bochum, Bochum, Germany • [8]High-Field and Hybrid MR Imaging, Essen University Hospital, University of Duisburg-Essen, Essen, Germany

eLife Assessment

This **important** study provides insights into the role of the cerebellum in fear conditioning, addressing a key gap in the literature. The evidence presented in support of the conclusions is **solid**. This work will be of interest to both the extinction learning and cerebellar research communities.

https://doi.org/10.7554/eLife.105399.2.sa2

Abstract

The key elements for fear extinction learning are unexpected omissions of expected aversive events, which are considered to be rewarding. Given its reception of reward information, we tested the hypothesis that the cerebellum contributes to reward-like prediction error processing driving extinction learning via its connections with the ventral tegmental area (VTA). Forty-three young and healthy participants performed a three-day fear conditioning paradigm in a 7T MR scanner. The cerebellum and VTA were active during unexpected omissions of aversive unconditioned stimuli in the initial extinction trials and in other learning phases, in line with the proposed role of prediction-error processing. Increased functional connectivity was observed between the cerebellum and VTA, indicating that they are functionally coupled during fear extinction learning. These results suggest that an interaction between the cerebellum and VTA should be incorporated into the existing model of the fear extinction network.

Introduction

Deficits in learning to extinguish previously associated threat responses to cues that no longer signal danger are thought to be one of the main causes in the development of anxiety disorders, including post-traumatic stress disorder and social anxiety disorder[1]. Exposure therapy addresses this deficit and attempts to compensate for failed or absent extinction learning from the past[2]. This is modeled effectively with extinction training in classical fear conditioning paradigms where the previously paired conditioned stimulus (CS+) is no longer followed by the aversive unconditioned stimulus (US)[3,4]. At the beginning of extinction training the omission of the US is unexpected and

this unexpected lack of the US is considered to be rewarding[5–7]. This reward-like prediction error is thought to drive safety learning. This newly learned safety association should then inhibit the initially learned fear association (which is never fully lost)[8].

The reward system in the brain is primarily associated with the mesolimbic dopamine system[9], i.e., the ventral tegmental area (VTA) and the ventral striatum. In fact, recent rodent studies provide strong evidence that the VTA is central for extinction learning, and that (reward-like) prediction errors are encoded in a subset of VTA neurons[5–7,10].

Importantly, new evidence indicates that the cerebellum receives reward signals[11–13], and has direct efferent connections with the mesolimbic dopaminergic system, in particular the VTA[14–17].

Functional magnetic resonance imaging (fMRI) findings in healthy human participants suggest that the cerebellum is involved in processing of prediction errors in fear conditioning[18,19]. Ernst et al.[18] studied acquisition of learned fear responses using a partial reinforcement rate, i.e., a subset of CS+ trials were not reinforced in acquisition training. Cerebellar activations were strongest in the unreinforced CS+ trials, when the US was expected but did not occur. An unreinforced CS+ trial can be considered as a very first extinction trial, and cerebellar activations may reflect prediction errors driving extinction learning. Likewise, recordings from the VTA in rodents showed strong activations at the time the US is expected and does not occur during initial extinction trials[6,10]. The aim of the present fMRI study is therefore to test the hypothesis that the cerebellum contributes to fear extinction learning via its connection to the VTA.

In contrast to the previous one-day study design[18], here we used a three-day design with acquisition training on day 1, extinction training on day 2 and a recall test on day 3 (Figure 1A). This multi-day study design allowed enough time for consolidation of both fear and extinction learning. Cerebellar activity related to the unexpected omission of the US was assessed in four different learning phases, that is, during extinction training (in initial CS+ trials), during the recall test (in initial CS+ trials), during reacquisition (in interspersed non-reinforced CS+ trials), and during reextinction (the initial CS+ trials immediately following reinforced trials) (Figure 1A). The CSs were geometric shapes (i.e. square and diamond), and the US was an electrical stimulation (Figure 1B). Full (100%) reinforcement was used during acquisition on day 1 to maximize the size of the reward-like prediction error elicited by the initial unexpected US omissions during extinction. Learning phases differed in the underlying learning state, ranging from newly acquired fear, to spontaneous recovery after consolidated extinction[20], to reactivated fear and reextinction, providing four distinct contexts in which to examine cerebellar and VTA responses to unexpected US omission[21].

We used a reinforcement-learning-based deep learning model fitted to group-averaged trial-by-trial skin conductance response (SCR) data to compute per-trial prediction errors. SCRs reflect autonomic arousal linked to stimulus salience and expectancy. Averaged SCRs were used here to capture predictive learning about the CS-US association at the population level, allowing us to derive model-based prediction error values. 7T fMRI allowed us to assess VTA activation.

Results showed activation in the cerebellum and VTA for all four phases related to unexpected US omissions, despite differences in learning contexts between phases. Results also showed significant functional interaction between the cerebellum and VTA related to prediction errors in unexpected omission trials. These findings suggest that the cerebellum contributes to extinction learning via the VTA, and thus supports safety learning inhibiting the initial fear association.

Results

We first report behavioral results demonstrating that participants acquired and partially extinguished conditioned fear in the three-day fear conditioning paradigm (Figure 1). We then present fMRI results related to the prediction, presentation and unexpected omission of the US in the cerebellar cortex, deep cerebellar nuclei (DCN), and VTA. Finally, we report analyses of functional connectivity between the cerebellum and the VTA.

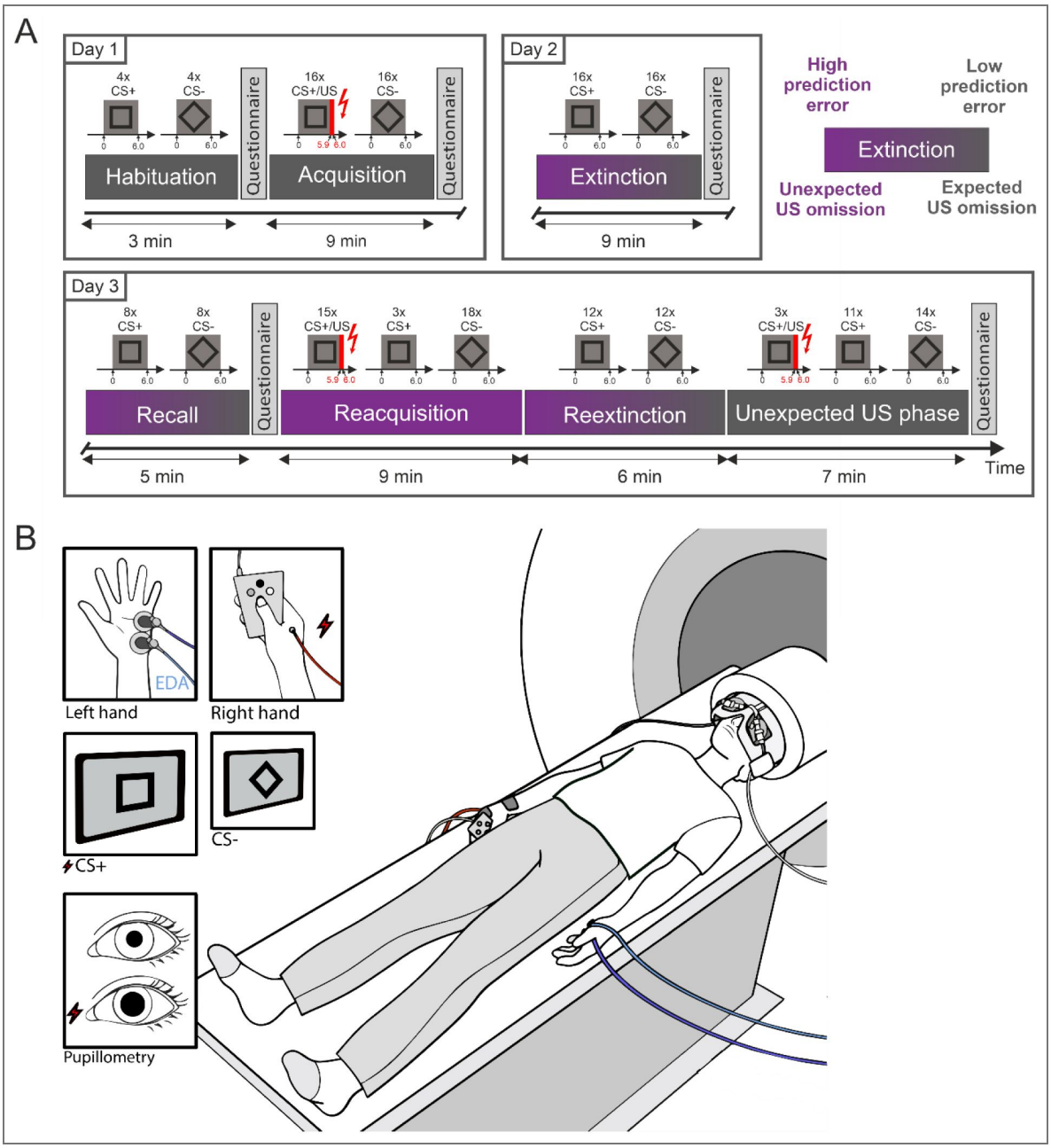

Figure 1.

A: Overview of the three-day fear conditioning paradigm. Day 1 consisted of habituation and acquisition training, day 2 of extinction training, and day 3 of a recall test followed by reacquisition, reextinction, and the unexpected US phase. Extinction, recall, reacquisition, and reextinction are highlighted in purple, indicating phases in which unexpected US omissions occurred. Extinction, recall, and reextinction are shown with a gradient to illustrate the gradual decrease in unexpectedness of US omissions across trials, in contrast to the discrete omission events during partially reinforced reacquisition. **B:** Schematic of the experimental setup during 7 T MRI. Participants lay supine in the scanner and viewed visual conditioned stimuli (CS; geometric shapes) displayed on a screen at the end of the bore via a mirror system. The eyes were simultaneously imaged with an eye-tracking system. Electrodermal activity (EDA) was recorded from the left hand (blue/purple cables). The unconditioned stimulus (US; electrical stimulation) was delivered to the right hand (red cable), and responses were recorded using a button box.

Behavioral data

Behavioral responses were assessed using skin conductance responses (SCRs; Figure 2), pupil size responses (PSRs; Figure 3), and self-reported ratings. Results for all three measures are reported below, with non-parametric ANOVA-type statistics and descriptive summaries of self-reported ratings provided in the Supplementary Information (Table S1-S6).

Skin conductance responses (SCRs)

During habituation (day 1), SCRs did not differ between CS+ and CS- (Figure 2, top row). SCRs for both stimuli decreased significantly across habituation trials. Non-parametric ANOVA-type statistics revealed a significant effect for Time (early vs. late; $F_1 = 26.49$; $p < 0.001$) but not for Stimulus (CS+ vs. CS-; $F_1 = 1.19$; $p = 0.276$) or the Stimulus x Time interaction ($F_1 = 0.06$; $p = 0.802$). During fear acquisition training (day 1), participants underwent a fully reinforced, semi-instructed differential fear conditioning protocol. Participants quickly learned the associations, and their responses decreased as expected due to habituation effects. SCRs for the CS+ were significantly higher than for the CS-, and decreased in late trials. Non-parametric ANOVA-type statistics revealed a significant effect of Stimulus (CS+ vs. CS-; $F_1 = 20.79$; $p < 0.001$) and Time (early vs. late; $F_1 = 28.75$; $p < 0.001$). No significant Stimulus x Time interaction was observed ($F_1 = 3.80$; $p = 0.051$).

During extinction training (day 2), SCRs were higher for the CS+ than for the CS- only in initial extinction trials. Non-parametric ANOVA-type statistics revealed a significant effect of Stimulus (CS+ vs. CS-; $F_1 = 7.71$; $p = 0.006$) and Time (early vs. late; $F_1 = 21.10$; $p < 0.001$), whereas the Stimulus x Time interaction was not significant ($F_1 = 3.74$; $p = 0.053$). Exploratory post hoc comparisons indicated higher SCRs to the CS+ than to the CS- during early extinction trials ($p = 0.017$), but no differentiation during late extinction ($p = 0.798$). These post hoc results are reported descriptively, as the Stimulus x Time interaction did not reach statistical significance.

During the recall test (day 3), SCRs for CS+ were significantly higher than for CS- (Figure 2, bottom row). In late recall, no difference between SCRs for CS+ and CS- was observed. Non-parametric ANOVA-type statistics revealed a significant main effect of Stimulus (CS+ vs. CS-; $F_1 = 4.99$; $p = 0.026$), Time (early vs. late; $F_1 = 36.74$; $p < 0.001$) and a Stimulus x Time interaction ($F_1 = 5.53$; $p = 0.019$). Post hoc tests revealed that the CS+ had a significantly higher SCR than the CS- in early trials ($p < 0.001$) but not in late trials ($p = 1.000$). SCRs for both stimuli significantly decreased in late trials compared to early trials (both: $p < 0.001$).

During reacquisition on day 3, SCRs for the CS+ were significantly higher than for the CS-, and decreased for both stimuli in late reacquisition. Non-parametric ANOVA-type statistics revealed a significant effect of Stimulus (CS+ vs. CS-; $F_1 = 27.20$, $p < 0.001$), Time (early vs. late; $F_1 = 35.44$, $p < 0.001$) and a Stimulus x Time interaction ($F_1 = 6.80$, $p = 0.009$). Post hoc tests showed that both stimuli were higher in early vs late reacquisition training (CS- early vs. CS-late; $p = 0.007$; CS+ early vs. CS+ late; $p < 0.001$), and a significant difference between stimulus types during early trials ($p < 0.001$), but no significant difference in late trials ($p = 0.094$).

During reextinction on day 3, SCRs related to the CS+ were significantly higher than for the CS- although the difference was small. Non-parametric ANOVA-type statistics revealed a significant main effect of Stimulus (CS+ vs. CS-; $F_1 = 4.47$; $p = 0.035$), but neither for Time (early vs. late; $F_1 = 0.08$; $p = 0.780$) nor for the Stimulus x Time interaction ($F_1 = 0.95$; $p = 0.329$).

In the last phase on day 3, three CS+ trials were unexpectedly reinforced (“unexpected US phase”). SCRs for the CS+ were significantly higher than for the CS-. Non-parametric ANOVA-type statistics revealed a significant main effect of Stimulus (CS+ vs. CS-; $F_1 = 5.43$; $p = 0.020$), but neither for Time (early vs. late; $F_1 = 0.29$; $p = 0.593$) nor for the Stimulus x Time interaction ($F_1 = 2.34$; $p = 0.126$).

Pupil size responses (PSRs)

During habituation on day 1, PSRs related to the CS+ and CS- did not differ significantly (non-parametric ANOVA-type statistics; all $p > 0.199$).

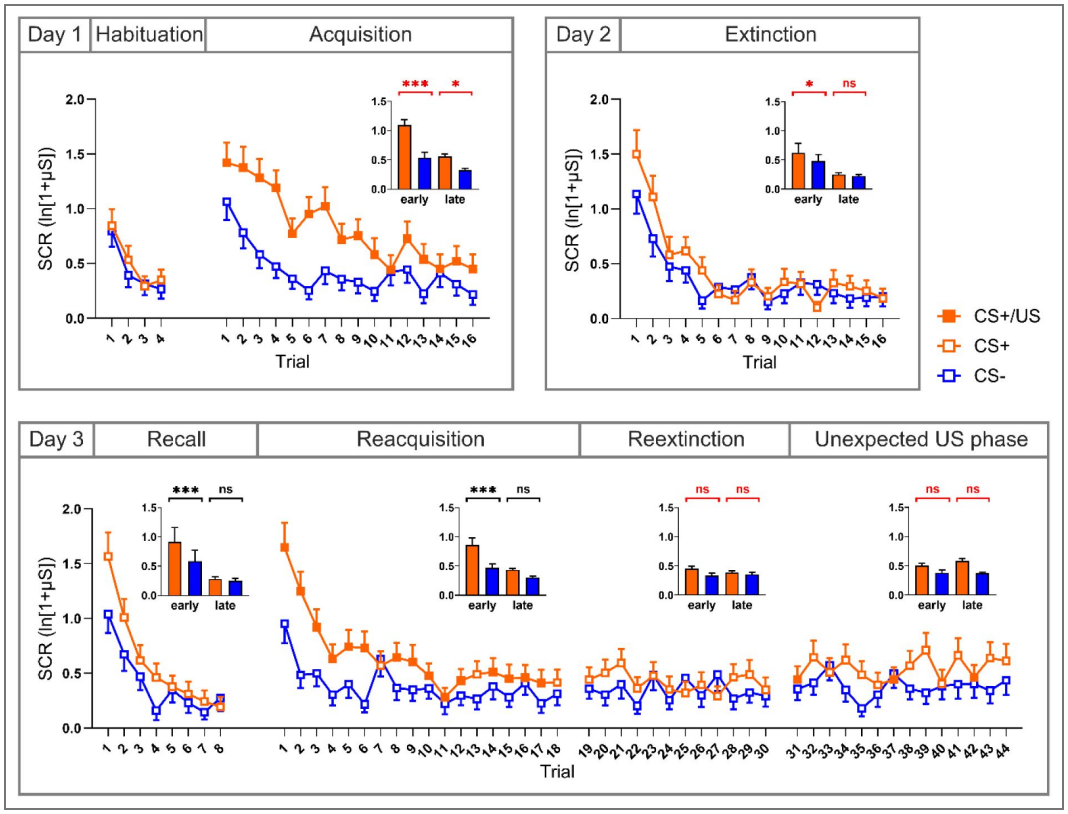

Figure 2. Skin conductance responses (SCRs) for each trial, with CS+ (shown in orange) and CS- (shown in blue) responses paired in blocks.

Reinforcement of the CS+ by a US (CS+/US) is indicated by filled squares. The CS- is never reinforced. Bar plots on the top right show mean SCRs for early and late trials, defined as the first and second halves of each phase, respectively. Significance markers indicate post hoc comparisons between CS+ and CS- within early or late trials; markers are shown in black when the Stimulus x Time interaction was significant and in red when the interaction was not significant. On day 1, there was no differentiation between CS+ and CS- in the habituation phase, with significant differentiation emerging during acquisition training. On day 2, differentiation was apparent in early extinction trials and was no longer evident later in extinction (trend-level effect; non-significant Stimulus x Time interaction). On day 3, during the initial recall test, participants exhibited renewed CS+/CS- differentiation, consistent with spontaneous recovery. During initial reacquisition, there were again differential responses to the CS+ and CS-, which appeared reduced in reextinction and the unexpected US phase. Mean values are shown with error bars representing the standard error of the mean. CS: Conditioned stimulus; US: Unconditioned stimulus; SCR: Skin conductance response; μS: microsiemens.

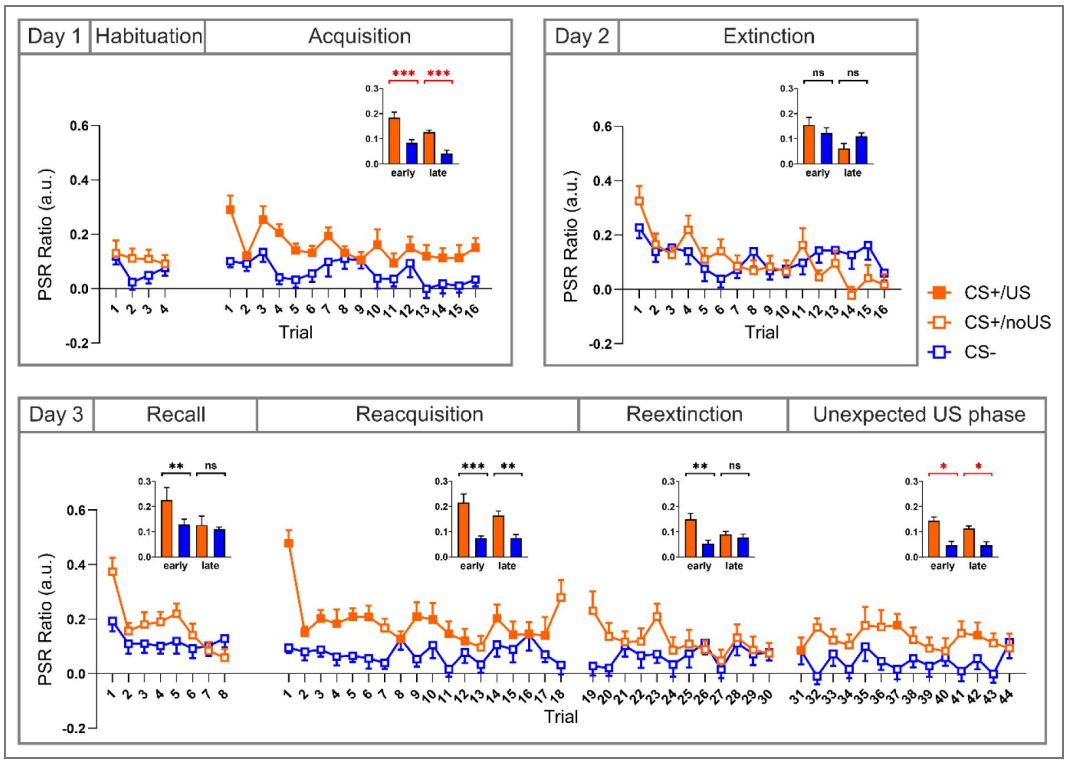

Figure 3. Pupil size responses (PSRs) for each trial, with CS+ (shown in orange) and CS- (shown in blue) responses paired in blocks.

Reinforcement of the CS+ by a US (CS+/US) is indicated by filled squares. The CS- is never reinforced. Bar plots on the top right show mean PSRs for early and late trials, defined as the first and second halves of each phase, respectively. Significance markers indicate post hoc comparisons between CS+ and CS- within early or late trials; markers are shown in black when the Stimulus x Time interaction was significant and in red when the interaction was not significant. On day 1, there was no differentiation between CS+ and CS- in the habituation phase, with significant differentiation emerging during acquisition training. On day 2, a significant Stimulus x Time interaction was observed during extinction, reflecting a decrease from early to late extinction trials for the CS+ but not for the CS-, despite the absence of significant post hoc CS+/CS- differences. On day 3, during the initial recall test, participants exhibited renewed CS+/CS- differentiation, consistent with spontaneous recovery. During initial reacquisition, there were again differential responses to the CS+ and CS-, which appeared reduced in reextinction and the unexpected US phase. Mean values are shown with error bars representing the standard error of the mean. CS: Conditioned stimulus; US: Unconditioned stimulus; PSR: Pupil size response.

During fear acquisition training on day 1, PSRs were significantly higher in CS+ trials than in CS- trials, and PSR was more pronounced in early compared to late trials (Figure 3, top row). Non-parametric ANOVA-type statistics revealed a significant main effect of Stimulus (CS+ vs. CS-; F_1 = 39.51; p < 0.001) and Time (early vs. late; F_1 = 24.31; p < 0.001). No significant Stimulus x Time interaction was observed (F_1 = 0.58; p = 0.446).

During extinction training on day 2, PSRs for the CS+ decreased from early to late extinction, but there were no differences between early and late trials for the CS-. Non-parametric ANOVA-type statistics revealed a significant main effect of Time (early vs. late; F_1 = 18.21; p < 0.001) and a significant Stimulus x Time interaction (F_1 = 4.28; p = 0.039). No significant main effect of Stimulus was observed (CS+ vs. CS-; F_1 = 1.12; p = 0.290). Post hoc tests showed that PSRs related to the CS+ were significantly lower during late extinction compared to early (p < 0.001). This result was not observed for PSRs related to the CS- (p = 0.643).

During the recall test on day 3, PSRs for the CS+ were significantly higher than for the CS-during early, but not late, recall trials (Figure 3, bottom row). Non-parametric ANOVA-type statistics revealed a significant effect of Stimulus (CS+ vs. CS-; F_1 = 6.59; p = 0.010) and Time (early vs. late; F_1 = 6.41; p = 0.011) and a significant Stimulus x Time interaction (F_1 = 4.87; p = 0.027). Post hoc tests showed higher PSRs for the CS+ than for the CS- during early recall (p = 0.017), but not during late recall (p = 0.862), and that the PSRs related to the late CS+ were significantly lower than early (p = 0.003)

During reacquisition on day 3, PSRs for the CS+ decreased from early to late trials, whereas no change over time was observed for the CS-. PSRs differed between CS+ and CS- in both early and late trials. Non-parametric ANOVA-type statistics revealed significant effects of Stimulus (CS+ vs. CS-; F_1 = 47.87, p < 0.001) and Time (early vs. late; F_1 = 4.54, p = 0.033), as well as a significant Stimulus x Time interaction (F_1 = 7.70, p = 0.006). Post hoc tests indicated that PSRs for the CS+ were higher early than late (p = 0.004), whereas no such change was observed for the CS- (p = 0.974).

During reextinction on day 3, PSRs for CS+ were significantly higher than for CS- early, but the difference became insignificant in late trials. Non-parametric ANOVA-type statistics revealed a significant main effect of Stimulus (CS+ vs. CS-; F_1 = 9.10; p = 0.003) and a Stimulus x Time interaction (F_1 = 4.43; p = 0.035), but not a significant main effect of Time (early vs. late; F_1 = 0.83, p = 0.364). Post hoc tests showed a significant difference between the CS+ and CS- early but not late (early: p = 0.007; late: p = 0.775).

During the unexpected US phase on day 3, PSRs for the CS+ were higher than for the CS-. Non-parametric ANOVA-type analysis revealed a significant main effect of Stimulus (CS+ vs. CS-; F_1 = 17.35; p < 0.001), but not of Time (early vs. late; F_1 = 0.83; p = 0.363) and the Stimulus x Time interaction (F_1 = 0.03; p = 0.872)

Self-reports

Valence, arousal, fear and US expectancy ratings of the CS+ did not differ significantly from the CS- before acquisition training (Table S5 and S6). Following acquisition training, the CS+ was rated as more arousing, fearful and unpleasant compared to the CS-, and participants expected the US to follow the CS+ but not the CS- (all p < 0.001). Although less pronounced, the CS+/CS- differentiation persisted after extinction training and recall test (all p < 0.004). However, extinction and recall CS+ ratings were significantly reduced compared to acquisition (all p ≤ 0.001), whereas CS- ratings remained unchanged (all p > 0.532), reflecting extinction learning. The difference between CS+ and CS- increased at the end of day 3, when CS+/US pairings were reintroduced.

Non-parametric ANOVA-type statistics showed a significant main effect of Stimulus, Time and a Stimulus x Time interaction (main effects and interaction; all p < 0.001; Table S6). Post hoc tests revealed significant differences between stimulus type (CS+, CS-) following acquisition training, extinction training, recall test and at the end of day 3 but not after habituation.

fMRI results

Analysis of BOLD fMRI responses revealed significant activations related to the prediction, the presentation and the unexpected omission of the US (Figure 4-7). Our primary focus was on activations and connectivity related to the unexpected omission of the US, which could be detected across four distinct phases: extinction training, recall test, reacquisition, and reextinction (Figure 6-8). Table S7 reports VOI-level summary statistics (mean contrast estimates, 95% confidence intervals, and Cohen's d) for a cerebellar cortex VOI derived from the conjunction of unexpected US omission contrasts, as well as for the deep cerebellar nuclei (DCN), and VTA across all fMRI contrasts, while Table S8-S12 report cluster-level activation and connectivity results.

Activations related to US prediction (CS+ > CS- and CS+ x prediction)

In the event-based analysis, there was no significantly higher activation in the cerebellum during the CS+ compared to the CS- (CS+ > CS-) during acquisition or extinction training. In the same contrast, the lateral VTA exerted significant differential activations during acquisition (Figure 4A; threshold free cluster enhancement (TFCE) and family-wise error (FWE) corrected $p < 0.05$), but not during extinction training (Figure 5A). Trend-level results showed differential cerebellar activation in posterolateral regions (lobule VI, Crus I, Crus II, lobule VIIb and VIIa; Figure 4A and Figure 4D; uncorrected $p < 0.05$) during acquisition training, which match results obtained by Ernst and colleagues (2019).

In the parametric modulation analysis, there were significant activations in the cerebellum (lobule and vermis I-VI), DCN and VTA negatively associated with CS+ prediction value modulation during acquisition training (CS+ x prediction; Figure 4B), whereas there were significant positively associated activations during extinction training (Figure 5B; lobule VI and Crus I bilaterally, right lobule V and VTA). Parametric modulation results indicate that early CS+ presentations are paired with high cerebellar activations in both acquisition and extinction training, with prediction values being low during early and high during late acquisition, whereas prediction values are high during early and low during late extinction (Figure 9). In other words, the activations were negative due to high activations of early CS+ compared to late CS+ events with prediction values taking the opposite course (i.e., low during early acquisition and high during late acquisition).

Activations related to US presentation (US post CS+ > no US post CS-)

During acquisition training, significant activations related to the presentation of the US (US post CS+ > no US post CS-) were widespread and included the whole cerebellum and VTA (Figure 4C). Cerebellar activations extended to the DCN. Peak activations in the cerebellum were found ipsilateral to the US (electrical stimulation applied to right hand).

Activations related to the unexpected omission of the US (first 3 no US post CS+ > no US post CS- and no US post CS+ x prediction error)

Considering the whole extinction training phase, trend-level results showed activations in posterolateral lobules and the VTA related to the omission of the US (no US post CS+ > no US post CS-; Figure 5C). When focusing specifically on the early trials in extinction training to quantify activations related to the unexpected omission of the US, we found significant activations in both event-based and parametric modulation analyses.

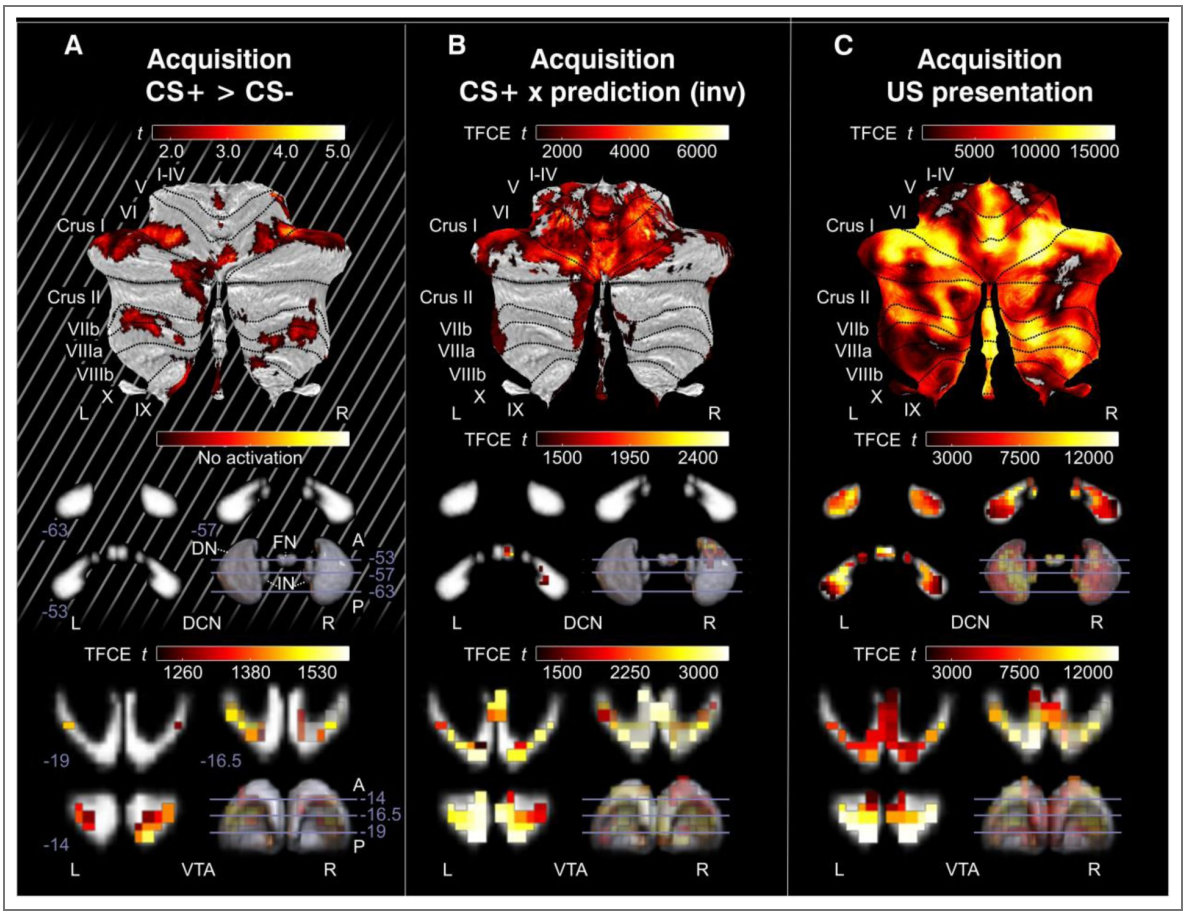

Figure 4. Cerebellar cortex, DCN, and VTA fMRI activations related to prediction and presentation of the US during acquisition training.

Cerebellar cortex activations are shown on cerebellar flatmaps (SUIT[22], while VTA and DCN activations are displayed on coronal slices progressing from posterior to anterior, with MNI y-coordinates indicated at the lower left of each slice. A 3D rendering in the bottom right shows slice locations within the probabilistic DCN and VTA atlases (see Methods for atlas generation). Trend-level results are indicated by gray diagonal lines in the background of the panels. **A:** The event-based CS+ > CS- prediction contrast showed trend-level activations in posterolateral cerebellar lobules and the VTA; no DCN activations were observed. **B:** Inverse activations were observed in the parametric modulation analysis of CS+ × prediction, with activations mainly in the anterior cerebellum, paravermis, lobule VI and Crus I, and the VTA, with trend-level activations in the DCN. **C:** Presentation of the US elicited widespread activations in the cerebellar cortex, DCN, and VTA. VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; CS: conditioned stimulus; US: unconditioned stimulus; MNI: Montreal Neurological Institute standard brain; SUIT: spatially unbiased atlas template of the cerebellum; TFCE t: threshold-free cluster-enhanced test statistic.

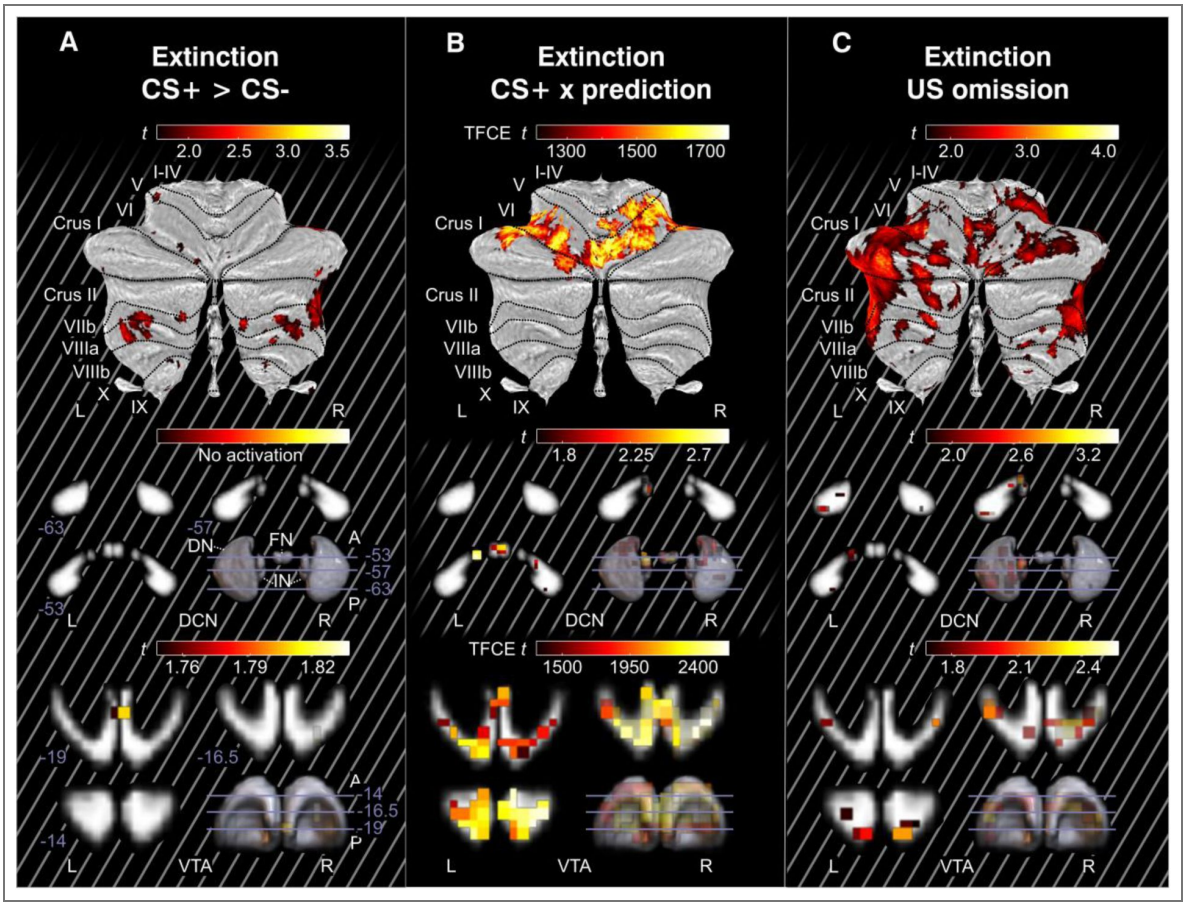

Figure 5. Cerebellar cortex, DCN and VTA fMRI activations related to the prediction and presentation of the US.

Cerebellar cortex activations are shown on cerebellar flatmaps (SUIT[22]), while VTA and DCN activations are displayed on coronal slices progressing from posterior to anterior with MNI y-coordinates indicated on the left-bottom for each slice. A 3D rendering in the bottom right shows the slice locations within the probabilistic DCN and VTA atlases (see methods for atlas generation). Trend-level results are indicated by gray diagonal lines in the background of the panels. **A:** The event-based CS+ > CS- contrast showed trend-level activations in lobules VIIb, VIIIa and VIIIb and the central VTA, with no DCN activations. **B:** The parametric modulation contrast CS+ x prediction showed activations in lobule VI and Crus I, the vermis, and the VTA, as well as trend-level activations in the DCN. **C:** Trend-level activations related to the omission of the US were observed in the cerebellar cortex, DCN and VTA. VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; CS: conditioned stimulus; US: unconditioned stimulus; MNI: Montreal Neurological Institute standard brain; SUIT: spatially unbiased atlas template of the cerebellum; TFCE t: threshold-free cluster-enhanced test statistic.

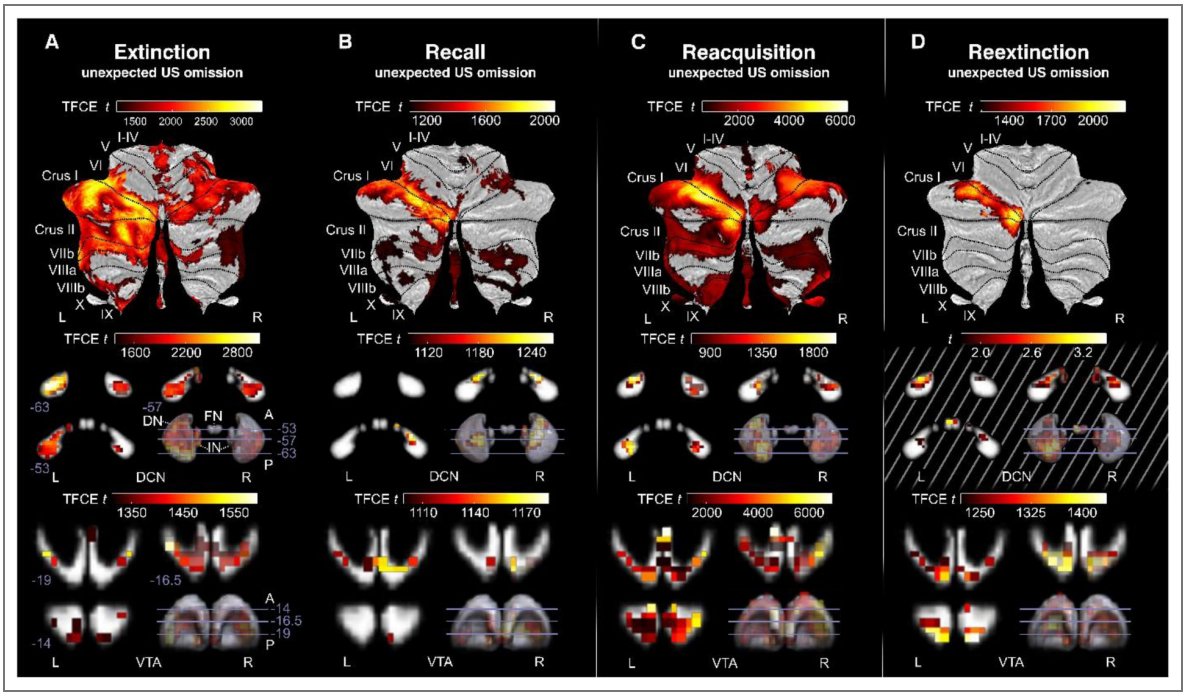

Figure 6. Cerebellar cortex, DCN and VTA fMRI activations to the unexpected omission of the US (event-based analysis).

Cerebellar cortex activations are shown on cerebellar flatmaps (SUIT[22]), while VTA and DCN activations are displayed in coronal slices progressing from posterior to anterior with MNI y-coordinates indicated on the left-bottom for each slice. A 3D rendering in the bottom right shows the slice locations within the probabilistic DCN and VTA atlas (see Methods for atlas generation). Trend-level results are indicated by gray diagonal lines in the background of the panels. Event-based contrasts of the first three unexpected omissions showed activations in the cerebellar cortex, DCN and VTA for all four phases. Activations in the DCN during reextinction were trend-level **(D)**. Activations were most consistent in left lobule VI and Crus I, with DCN activation mainly in the left dentate, while VTA activations were bilateral. VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; CS: conditioned stimulus; US: unconditioned stimulus; MNI: Montreal Neurological Institute standard brain; SUIT: spatially unbiased atlas template of the cerebellum; TFCE t: threshold-free cluster-enhanced test statistic.

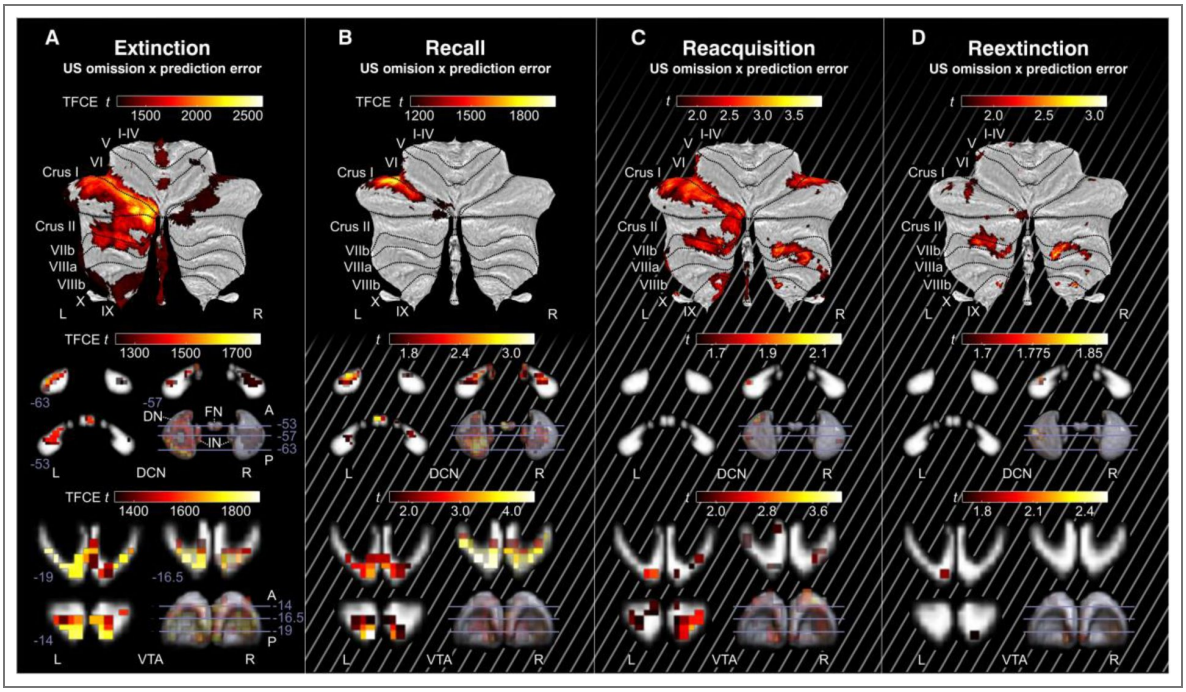

Figure 7. Cerebellar cortex, DCN and VTA fMRI activations to the unexpected omission of the US.

Cerebellar cortex activations are shown on cerebellar flatmaps (SUIT[22]), while VTA and DCN activations are displayed in coronal slices progressing from posterior to anterior with MNI y-coordinates indicated on the left-bottom for each slice. A 3D rendering in the bottom right shows the slice locations within the probabilistic DCN and VTA atlas (see methods for atlas generation). Trend-level results are indicated by gray diagonal lines in the background of the panels. Parametric modulation contrasts related to prediction errors from unexpected US omissions showed activations in all four phases in the cerebellum, DCN and VTA. Activations during the recall test in the DCN and VTA were trend-level (B). Activations in reacquisition and reextinction were trend-level (C, D). Activations were most consistently found in lobule VI and Crus I. In addition, activations were found in the vermis, Crus II and lobule VIIb. VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; CS: conditioned stimulus; US: unconditioned stimulus; MNI: Montreal Neurological Institute standard brain; SUIT: Spatially Unbiased Atlas Template of the cerebellum; TFCE t: threshold-free cluster-enhanced test-statistic.

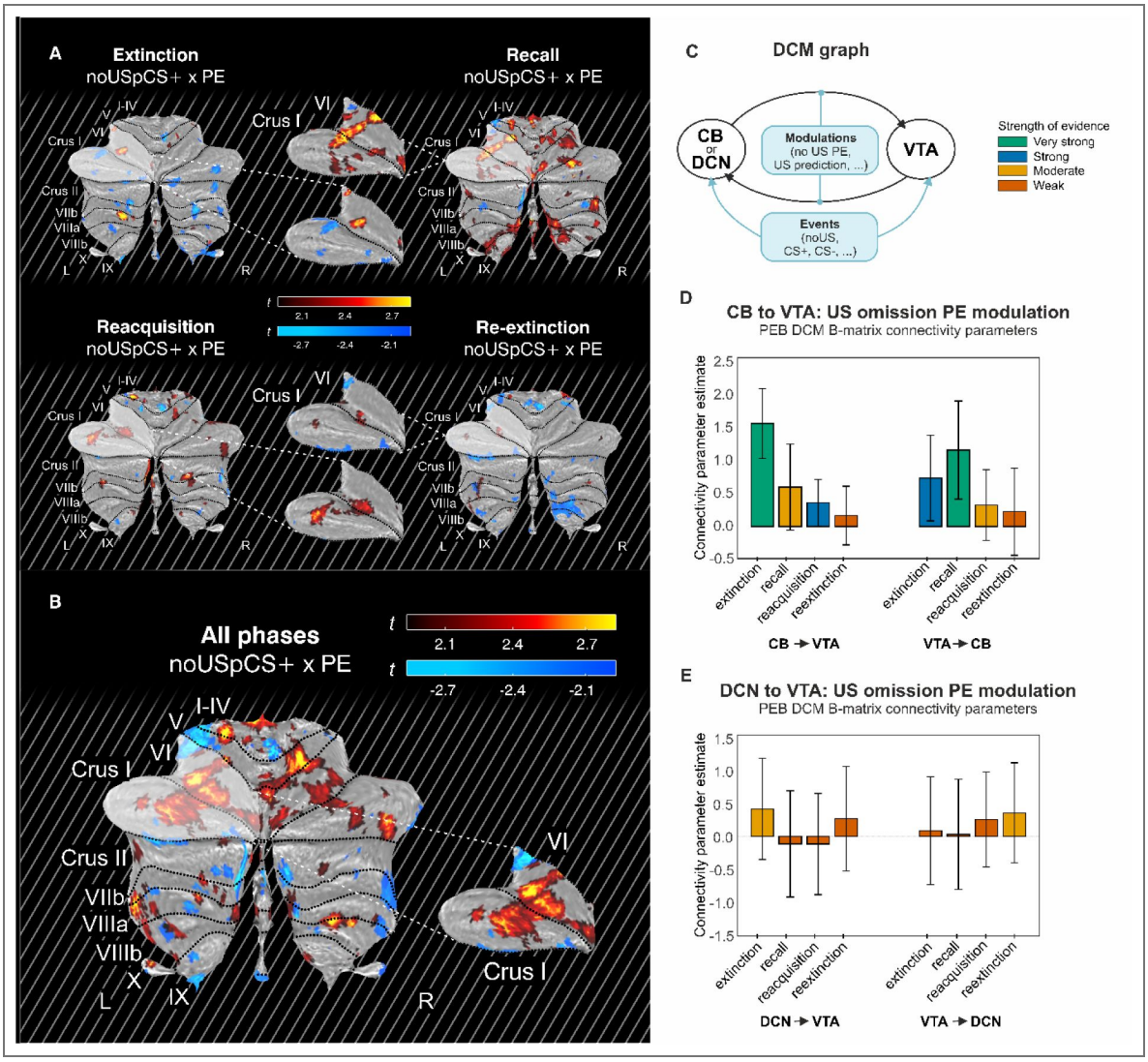

Figure 8. Functional connectivity analysis using PPI and DCM.

A,B: Trend-level activations in the cerebellum found by PPI analysis using the VTA as a seed region, shown on cerebellar flatmaps[22]. To highlight activations, a zoomed cutout of lobule VI and Crus I is also shown. Trend-level results are indicated by gray diagonal lines in the background of the panels. **A:** Trend-level connectivity with VTA is found in lobule VI and Crus I for extinction training, the recall test, reacquisition and reextinction using PPI of the unexpected US omission contrasts from the parametric modulation analysis with a VTA seed. **B:** Summary of trend-level PPI results by summing all contrasts shown in panel A. **C-E:** DCM analysis showed significant modulation of cerebellar cortex (CB) to VTA connections with prediction errors during unexpected omission events. **C:** DCM model is shown, modulations act on both the CB/DCN to VTA as well as the VTA to CB/DCN connections. Events, i.e., US post CS+, no US post CS+, no US post CS-, CS+ and CS- for each phase, are provided as inputs to both nodes. Colors of bars are determined by the posterior probability calculated in the PEB analysis (very strong: P > 0.99; strong: 0.95 < P < 0.99; moderate: 0.75<P < 0.95; weak: P < 0.75). Very strong and strong results are considered significant, moderate results are considered trend-level. **D:** Significant modulations during extinction training, the recall test and reacquisition. The strongest result was found during extinction training in the connection from the cerebellum to VTA. **E:** Trend-level results showed connectivity from the DCN to VTA in extinction, and from the VTA to DCN in reextinction. PPI: Psychophysiological interaction; DCM: Dynamic causal modeling; PEB: Parametric Empirical Bayes; PE: Prediction error; P: Posterior probability; VTA: Ventral tegmental area; CB: Cerebellar region in lobule VI and Crus I consistently active during unexpected US omissions; US: unconditioned stimulus; L: left; R: right; SUIT: spatially unbiased atlas template of the cerebellum; t: test-statistic.

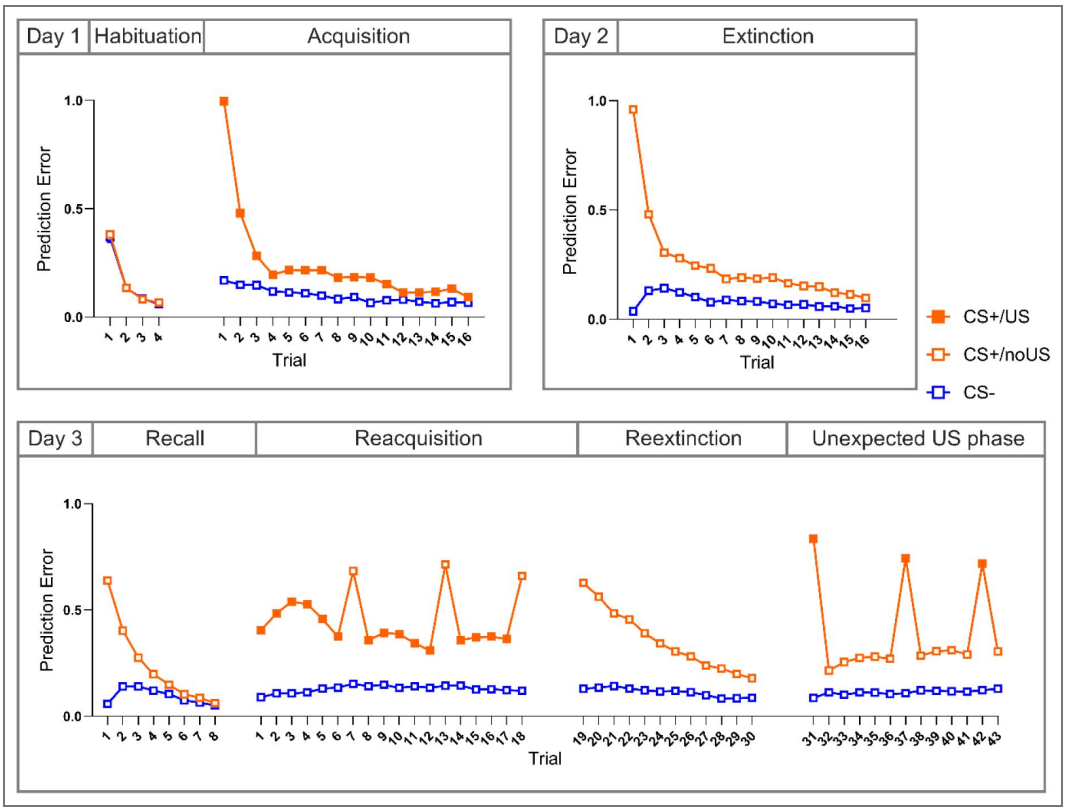

Figure 9. Mean absolute prediction error values for each trial estimated from SCRs, with CS+ and CS- responses paired in blocks.

Reinforcement of the CS+ by a US (CS+/US) is indicated by filled squares. Prediction and prediction error values were estimated over 200 iterations. CS: conditioned stimulus; US: unconditioned stimulus.

In the event-based analysis, the cerebellar cortex, DCN and VTA showed significant activations related to unexpected omissions of the US in initial trials for all four phases (first 3 no US post CS+ > no US post CS-; extinction training, recall test, reacquisition and reextinction; Figure 6). The location of activations was consistent across phases, with highest activations in left posterolateral lobule VI and Crus I. Activations in the left posterolateral cerebellar cortex were paired with high activations in the left dentate nucleus. The VTA was bilaterally activated.

In the parametric modulation analysis, significant activations were found in the cerebellum for unexpected omissions during extinction training and the recall test (Figure 7A and Figure 7B), and showed trend-level activations during reacquisition and reextinction (Figure 7C and Figure 7D). The highest activations were again found in left lobule VI and Crus I. In the DCN, significant activations were found during extinction training with peak activations in the left dentate, with trend-level results in the recall test. In the VTA, results were significant during extinction training (Figure 7A), with trend-level activations during the recall test, reacquisition and reextinction.

PPI: Functional connectivity during unexpected US omissions between the cerebellum and VTA

Next, we aimed to assess functional interactions between the cerebellum and VTA using psychophysiological interaction (PPI), using the VTA as a seed region. In the parametric modulation analysis, we observed trend-level positive functional connectivity in left lobule VI and Crus I for extinction training, recall test, reacquisition and reextinction (Figure 8A). When combining the contrasts by summing, the combination of all four phases showed similar positive functional connectivity between the cerebellar cortex and the VTA in lobule VI and Crus I, but only on a trend-level (Figure 8B). No activations were seen in the DCN.

To assess whether similar connectivity patterns were observed when using an alternative seed, we performed an additional PPI analysis using a cerebellar region of interest defined from the conjunction analysis of unexpected US omissions (see Volumes of interest (VOI) definition in Methods and Figure S4) as the seed. In this analysis, connectivity effects in the VTA were very limited and restricted to a small number of voxels at an uncorrected threshold (Figure S3).

DCM: Functional connectivity during unexpected US omissions between the cerebellum and VTA

In the dynamic causal modeling (DCM) parametric modulation analysis, unexpected omissions significantly positively modulated both the connection from cerebellar cortex to VTA and VTA to cerebellar cortex during extinction training (Figure 8D; very strong evidence, posterior probability $P > 0.99$, strong evidence: $0.95 < P < 0.99$). In the recall test, the VTA to cerebellar cortex connection was positively modulated. In reacquisition, the cerebellar cortex to VTA connection was positively modulated. Results were less strong for the other unexpected omission modulations, with trend-level moderate evidence for the cerebellar cortex (CB) to VTA connection in the recall test and the VTA to CB connections in reacquisition, and weak evidence for both CB to VTA and VTA to CB connections in reextinction (Figure 8D; moderate $0.75 < P < 0.95$; weak $P < 0.75$). In the model incorporating the DCN and VTA, there was a trend-level modulation from the DCN to the VTA during extinction training, as well as a trend-level modulation from the VTA to the DCN during reextinction (Figure 8E; moderate significance, $0.75 < P < 0.95$).

Discussion

The cerebellum and VTA were active during unexpected omissions of aversive unconditioned stimuli in the initial extinction training trials and in other learning phases. Notably, this activation was accompanied by increased functional connectivity between the cerebellum and VTA. Given that the unexpected omission of an aversive stimulus can be considered rewarding, our data suggest that interactions between the cerebellum and the VTA contribute to reward-like prediction error processing during fear extinction learning in humans. Such interactions may represent one

mechanism by which the cerebellum contributes to non-motor functions, in this case the control of emotions. Our findings build on previous research in rodents showing that the cerebellum modulates addictive, social and depression-like behavior via direct projection to VTA[16,23].

Cerebellar activations were most prominent in trials with maximum prediction errors, i.e., during the initial extinction training phase, but were also seen in early recall, reacquisition and early reextinction. The activations in lobule VI and Crus I are consistent with previous findings by Ernst et al.[18] during unexpected omissions in a partially reinforced acquisition phase.

Additionally, other studies have reported activations in these regions during the prediction of the US in extinction training[24–27] and during unexpected omissions of aversive stimuli[28–31]. More broadly, activations in lobule VI and Crus I have been observed during US prediction[32], particularly in the early phases of acquisition training. Conversely, unexpected rewards lead to activation in Crus I[33], and a meta-analysis implicated left lobule VI and Crus I in both the anticipation and processing of rewarding outcomes[34]. Anticipation-related activations were more hemispheric, whereas rewarding outcome activations were more paravermal.

Activated areas in more lateral parts of lobule VI and Crus I overlap with multi-demand regions[35], whereas activations neighboring the vermis overlap with emotional regions[36]. The main components of the multi-demand (or executive) neural network are working memory, attention and inhibition[37]. Thus, attention-related processes involved in fear conditioning may also contribute to cerebellar activations. Attention-related processes have been related to the salience of the stimulus[38], but are also thought to change when a cue is followed by an unexpected outcome with more attention being paid in the next trial[39,40]. Attention related processes may therefore contribute to activations in the more lateral cerebellar parts of lobule VI/Crus I. However, unexpected omission of the US should lead to increased attention to the CS in the next trial and therefore increased activation towards the CS (which was not the case) and not at the time the US was presented and did not occur.

Our findings of increased activation in the VTA related to the unexpected omission of the aversive US align well with rodent data[6,7]. In rodents, it has been shown that dopamine neurons in the VTA are active in initial extinction trials, and that lack of this dopamine signal prevents extinction learning[5,6,41]. In a more recent study, Salinas-Hernandez et al.[10] observed that these dopamine neurons project to extinction-related neurons of the nucleus accumbens, and receive input from the dorsal raphe. Our data show that the VTA is also involved in extinction learning in humans, and provide first evidence that the VTA is functionally coupled with the cerebellum during extinction learning. Esser et al.[42] found that the administration of L-Dopa in healthy human participants enhanced activation of the nucleus accumbens which was associated with VTA activity at the time of unexpected omission of the US in early extinction learning, which is in good agreement with our findings.

The unexpected omission of the aversive US is thought to be rewarding and fMRI signals in the cerebellum and VTA may be related to reward-like prediction errors. Whereas the VTA has long been known to process rewards, and fMRI data in humans indicate that the VTA is active related to the unexpected presentation of rewards[43,44], the cerebellum has only recently been found to process reward signals in rodents[11,13,45]. fMRI signals are known to be driven by synaptic input[46–48]. In the cerebellum this is mainly related to mossy fiber input to the cerebellar cortex[49]. There is evidence, however, that granule cell activity also modulates vasodilatation and likely contributes to fMRI BOLD signals in the cerebellar cortex[50,51]. Likewise, recent evidence suggests that activation of dopamine neurons in the VTA leads to increased BOLD signals using DREADD-fMRI in rodents[52]. Human fMRI-PET studies have shown that VTA BOLD activation in response to rewards correlates with increases in dopamine in the striatum, suggesting that the fMRI signal in the SN/VTA is related to dopaminergic neuron activity[44,53]. fMRI signals in the cerebellar cortex and VTA may therefore reflect incoming reward-related signals, prediction error processing or, maybe most likely, both.

The origin of the afferent reward signals to the cerebellum remains unclear; while an early study showed afferent connections from the VTA to the cerebellum[54], this finding has not been replicated in more recent research[55]. However, preliminary evidence from Guarque-Chabrera et al.[56] suggests the existence of a connection from the VTA to the cerebellum, although its strength and functional relevance remain uncertain. Given this limited evidence, the VTA-to-cerebellum interactions detected with DCM may not be driven primarily by a direct connection. Instead, they may reflect indirect network-level interactions within the mesolimbic and mesocortical systems.

In rodents, a subset of VTA dopamine neurons has been found to be active when the US is expected and does not occur, and are therefore considered to represent (reward-like) prediction error signals driving extinction learning[6,10]. The VTA, however, also contains neurons which are active during the presentation and prediction of threat[57,58], and likewise we observed fMRI activation of the VTA related to threat prediction and presentation. Cerebellar activations and their interactions are likely not limited to reward-like prediction errors; they may also be related to unexpected presentation of aversive stimuli. Because aversive stimulus presentation results in pronounced cerebellar activations and the response to the US habituates over time, we were unable to separate cerebellar activation related to the unexpected (initial acquisition trials) and the expected (late acquisition trials) presentation of the US.

Although the present findings are consistent with the interpretation of prediction error related activations in the VTA (and cerebellum), based on the present fMRI data, we cannot decide whether activations reflect surprise signals (i.e., unsigned prediction error related signals) or signed prediction error signals[39,59–61]. Rodent data show that cerebellar neurons in the cortex and nuclei increase activations both related to the unexpected presentation and omission of reward signals[62], which would indicate that the cerebellum primarily contributes to unsigned (reward) prediction errors. Likewise, a meta-analysis of human prediction error showed that activations in the cerebellum were more consistently related to unsigned prediction errors[63].

Results showed significant functional interaction between the cerebellum and VTA during unexpected US omission in extinction training, and interactions were also observed during spontaneous recovery and reacquisition. This suggests that cerebellum-VTA coupling consistently emerges in conditions where prediction errors occur. Predictive signals of the cerebellum may be linked to the timing of the unconditioned response and therefore its omission and may interact with VTA-related prediction error processing.

Although interpretation must remain cautious given BOLD and DCM limitations outlined in more detail below, we detected phase-specific differences in cerebellum-VTA connectivity. Extinction learning showed bidirectional coupling, spontaneous recovery showed primarily VTA-to-cerebellum influences, and reacquisition showed predominantly cerebellum-to-VTA influences. A tentative interpretation is that a bidirectional loop may support the formation of an extinction memory during active learning, whereas in spontaneous recovery the extinction memory is mainly retrieved, possibly reducing the need for cerebellar output. In reacquisition, reactivation of the fear memory may again engage cerebellar outputs that modulate omission-related processing in the VTA, while the reactivated fear memory during reinforced trials may reduce the detectability of VTA-to-cerebellum influences.

Notably, the unexpected US omission area in the cerebellum was more prominent on the left, while we applied the US to the right. In our previous work it was also the left cerebellum which was mainly activated, however, stimulation was always done on the left[18,64]. As outlined above, we believe that cerebellar activation is reward related. In a recent meta-analysis of fMRI data in monetary reward tasks, reward outcome was also related to the left cerebellum (i.e., lobule VI in the cerebellar hemisphere and vermis)[34]. There is good evidence that emotional information is most prominently processed in the right cerebral hemisphere[65,66]. The present data suggest that emotional processing may also be lateralized in the cerebellum, with the left cerebellar hemisphere projecting to the right cerebral cortex and vice versa[67]. Lateralization of positive emotions to left cerebellar lobule VI has also been reported by Liu et al.[68]. Furthermore, emotion-related tasks resulted in increased activity of the left cerebellar hemisphere in patients with anorexia nervosa compared to controls[69]. Meta-analyses of fMRI studies of emotional tasks,

however, report mostly bilateral cerebellar activations with no clear lateralization[70–72]. These studies primarily focused on emotion recognition tasks. Future research should explore the lateralization of emotional processing in the cerebellum using a variety of emotional tasks. In addition to emotions, there is also a preference of the right cerebral hemisphere for the ventral fronto-parietal attention network, which has been related to the detection of potentially harmful events[73]. Thus, as outlined above, attention-related processes involved in fear conditioning may also contribute to (lateralized) cerebellar activations.

In the following, we discuss potential limitations of our study. Firstly, while we demonstrated functional connectivity in the DCM analysis, the results were weaker with PPI. PPI is known to frequently lack power and to have a high proportion of false negatives[74,75], particularly for event-based designs. Additionally, the low number of unexpected events in our experiment, due to the 100% reinforcement/non-reinforcement rates in acquisition training (which was designed to induce maximum prediction error in the initial extinction trials), may have made detecting activation challenging. Furthermore, as the PPI regressor is demeaned, any persistent activation (e.g., due to an anatomical connection) will decrease detected connectivity. Because cerebellar and VTA activation co-occur not only in response to unexpected US omissions, but also to US presentation and prediction, this lack of specificity could have weakened our results. The same reasons apply for the lack of observation in functional connectivity between the DCN and the VTA.

Secondly, in the DCM analysis, we observed strong modulation related to unexpected US omissions when using the cerebellar cortex and VTA as volumes of interest (VOIs), but observed only trend-level results when substituting the cerebellar cortex with the DCN. This is likely explained due to the high iron content in the DCN causing susceptibility artefacts and reducing signal[76]. Although we opted for a whole-brain sequence with a conventional echo-time in order to generate a dataset of general value, future studies could opt for lower echo times and optimize for detection of signal in the DCN[77].

Thirdly, the extent of the VTA varies depending on the MRI atlas used. The VTA in the atlas from Pauli et al.[78] is about 4 times smaller than the VTA in the atlas by Trutti et al.[79] due to different VTA definitions used. Our average volumes are in good accordance with Trutti et al.[79], which in turn is in good accordance with histological data. The VTA and its boundaries, particularly laterally, are not clearly visible on MRI scans and were identified by using landmarks such as the substantia nigra, red nucleus, interpeduncular fossa and the cerebral aqueduct. We have approximated the VTA region as closely as we could, but cannot rule out imperfect masks, especially at both lateral ends of the VTA.

In conclusion, our study provides evidence that interactions between the cerebellum and VTA contribute to fear extinction learning. These interactions may support reward-like prediction error processing, a mechanism thought to drive fear extinction learning. Previous studies found that the cerebellum modulates addictive and social behavior via direct projection to VTA in rodents[16,23]. In humans, our findings extend this framework by highlighting cerebellum-VTA interactions involved in emotional processing. Incorporating these interactions into models of the fear extinction network may provide new ways of improving exposure therapy by targeting the dopaminergic system or stimulating the cerebellum specifically during unexpected learning events.

Methods

Preregistration

This study was preregistered on OSF on 16/11/2023[80]. The preregistration document can be viewed at *https://doi.org/10.17605/OSF.IO/2PXWE*.

The preregistration outlines the methods explained here, and includes the main hypotheses. Adjustments were made to specific parts of the methods, which include the optimized localization of the VTA and DCN by manual drawing by an expert annotator, and the inclusion as volumes of interest (VOIs) in the functional connectivity analyses.

Deviations from preregistration

A primary deviation from the preregistration involved the event-based fMRI analysis of unexpected US omissions. The preregistered analysis plan specified event-based contrasts using early versus late halves of extinction, recall, reacquisition, reextinction and unexpected US phases. In addition, the preregistration described an exploratory analysis focusing on the initial three unexpected omission trials within these phases. In the final analyses, this preregistered exploratory first-three-trial contrast was used as the main event-based fMRI analysis, replacing the early/late block contrasts. This decision was supported by the rapid decay of behavioral differentiation and model-derived prediction error estimates observed under the full (non-)reinforcement structure of the paradigm (100% reinforcement during acquisition and 0% during extinction and recall), which indicated that prediction errors were largely confined to the initial trials (see Figure 2 for SCRs, Figure 3 for PSRs, and Figure 9 for prediction error estimates).

Furthermore, several preregistered exploratory analyses were not performed. In particular, a preregistered exploratory multi-node DCM of the broader fear extinction network was omitted as it was out of scope for the current VOI-specific manuscript centered on cerebellum-VTA interactions. Other preregistered exploratory analyses were likewise not performed, as they were not necessary to address the primary hypotheses of the present study.

In addition, the preregistration contains a reporting error in the sample size rationale: the alpha level was listed as $p < 0.05$, whereas the underlying power analysis conducted during grant preparation used $p < 0.005$. This affects the estimated minimum sample size but not the target sample size, which was 50 participants in both the preregistration and the original manuscript. The correct parameters are reported below.

Participants

Experiment power estimate and number of participants

Experiment power was estimated based on previous fear conditioning data acquired at 7T fMRI[18] using the FMRIPower toolbox (*fmripower.org*)[81]. Considering first level unexpected US omission contrasts during a partially reinforced acquisition training and aiming for a power of 80% at $p < 0.005$ for a one-sided hypothesis test, group sizes were estimated at 41 participants for Crus I ipsilateral to US presentation (effect size 0.56 in units of standard deviations). To account for potential dropouts and outliers, 50 participants were initially recruited. However, only 44 successfully completed the 3-day study, with reasons for non-completion including Covid-related issues (1 participant), problems with the Digitimer equipment (3 participants), and failure to attend on day 3 (2 participants). Additionally, one participant dropped out due to a shimming failure, resulting in a final sample size of 43 participants.

Exclusion criteria and demographic information

Participants were between 20 and 35 years old (23 men, 20 women, mean age: 24.7 (SD = 3.4) years), fluent in German, non-smokers, reported no intake of medication or illicit drugs affecting the central nervous system, had no history of neurological or mental disorders for themselves or their first-degree relatives, and had no previous participation in similar learning experiments. Additionally, female participants were not pregnant, breastfeeding, or using hormonal contraceptives. Before the experiment, participants completed the Depression Anxiety Stress Scale (DASS21G)[82]. Participants who scored higher than 20 for depression, 14 for anxiety or 25 for stress components were excluded from the study. All participants were right-handed based on the Edinburgh handedness inventory[83]. Participants were asked to refrain from alcohol consumption the night before the experiment. Informed consent was obtained from all participants. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki.

Fear conditioning paradigm

Participants underwent a three-day differential fear conditioning paradigm. The paradigm presentation was controlled by a computer running Presentation (version 21.0, Neurobehavioral System Inc, Berkeley, CA). Participants were shown two visual stimuli (square and diamond) as CS (CS+ and CS-) and informed that electrical stimulations would be applied during the experiment. Three types of trials were used: reinforced CS+ (CS+ paired with the US; CS+/US), unreinforced CS+ (no US) and unreinforced CS-. The CS- was never reinforced.

The paradigm consisted of seven phases separated over three days (Figure 1A). On day 1, the experiment started with a habituation phase consisting of 8 unreinforced CS- and CS+ trials (4 CS+ trials, 4 CS- trials, presented in semi-randomized order). This phase is intended to decrease novelty effects and enable participants to acclimate to the scanner environment and presentation of visual stimuli. The habituation phase was followed by acquisition training consisting of 32 trials with fully reinforced CS+ (16 CS+/US, 16 CS- trials), during which the fear association was learned. On day 2, extinction training was applied and consisted of 32 unreinforced trials (16 CS+, 16 CS-). On day 3, the recall test, reacquisition, reextinction and the unexpected US phase were applied. The recall test consisted of 16 unreinforced trials (8 CS+, 8 CS-). This phase tested the recall of extinguished fear responses and showed spontaneous recovery (see Results). Reacquisition, reextinction, and the unexpected US phases were done in a single continuous fMRI run without pauses between sequences. The partially reinforced reacquisition phases consisted of 36 trials (15 CS+/US trials, 3 CS+, 18 CS-) with 3 unexpected omissions of the US, reextinction consisted of 24 trials (12 CS+, 12 CS-) and the unexpected US phase consisted of 28 trials (3 CS+/US trials, 11 CS+, 14 CS-) with 3 unexpected reinforced CS+ trials (Figure 1A). Learned fear responses were measured by skin conductance responses (SCRs) and pupil size reactions (PSRs) in each trial. The reacquisition, reextinction and unexpected US phases were included to increase the number of trials with high prediction errors (i.e., discrepancy between US expectation and actual US presentation).

We used the model from Batsikadze et al.[30] to optimize our experimental design. Values for US prediction and US prediction error produced by the model were recorded during simulation. Simulations were repeated for different paradigm CS sequences, and the sequence which required a short number of trials and still yielded high prediction errors was selected.

The CS+ and CS- were presented in equal numbers during the first and second half of each phase, regardless of whether the CS+ was followed by a US or not. Counterbalancing was implemented for the visual CS (CS+ square and CS- diamond, or CS+ diamond and CS- square) as well as for the order of presentation of the CS during the unreinforced habituation, extinction training, and recall test phases (i.e., CS+ presented first or CS- presented first). The reinforced acquisition training and reacquisition phases always started with a CS+. The two orders of CS presentation were pseudorandomly generated, similar to Ernst et al.[18]. Participants were instructed to pay attention to any possible connection between the CS and US presentations. They were also informed that the patterns of CS and US presentations would stay the same across all phases. However, they were not told about the CS/US contingencies or the timing and occurrence of the US.

In the analysis, the focus was on trials containing unexpected omissions of the US (CS+/no US) with high prediction error values. These trials occurred during early extinction, early recall, reacquisition and the reextinction phase.

Visual conditioned stimuli (CS)

Visual stimuli were displayed on an fMRI monitor (BOLD-screen 32, Cambridge Research Systems Ltd., Rochester, UK) placed at the end of the scanner bore which was projected to the participant using a mirror system (Figure 1B). Two pictures of black geometric figures (a square and a diamond shape) on a gray background were used as CS+ and CS-. Both CS were presented for 6 s. The period of the CS was shortened from 8 s used in our previous study[18] to 6 s to decrease the length of the experiment, while still being able to assess SCRs[4]. In the periods between visual stimuli a fixation cross was shown (time of ITI: varies between 9 s and 13.5 s).

Aversive unconditioned stimulus (US)

In reinforced CS+ trials, the visual stimulus co-terminated with the presentation of the aversive US, an electrical stimulation (100 ms) applied to the first dorsal interosseous muscle area of the right hand via a 6.5 mm concentric surface electrode (WASP electrode, Specialty Developments, Bexley, UK). The US was produced by a high voltage DC stimulator (DS7A, Digitimer Ltd., London, UK) and consisted of four consecutive 500 µs current pulses (maximum output voltage: 400 V) with an interpulse interval of 33 ms. The location of the US electrode was kept constant across all days of measurement by marking the electrode position with a permanent marker on day 1 and 2. Before the first experimental phase began, the individual electrical stimulation intensity threshold was calibrated to be perceived as 'very unpleasant but not painful'. To reach the threshold, the current strength was progressively increased and modulated according to each participant's feedback. The calibrated US intensity was increased by 20 % to compensate for possible habituation to the US, which could result in a weakening of the conditioned responses as successfully done in previous studies of our group[30,84]. The US intensity remained the same for the three days of the experiment (mean US intensity: 3.8 (SD = 2.9) mA, range 0.6-19.2 mA).

Physiological data acquisition

During each phase, SCRs and pupil size were recorded. SCRs were acquired with appropriate hardware filters sampling at 2 kHz through an MP160 Data Acquisition Hardware unit (BIOPAC Systems Inc, Goleta, CA). Two SC electrodes were attached to the left-hand hypothenar eminence. Pupil size was measured using an MRI compatible eye tracking system (Avotec Inc, Stuart, FL). Calibration of the eye tracking system was performed prior to each phase to track gaze position on the screen.

Self-reports

After each learning phase, participants completed questionnaires using a 4-button fiberoptic response device (Current Designs, Haverford, PA). The Likert scale questionnaires assessed self-reports such as arousal (rated on a 1-9 scale from 1: "very calm" to 9: "very nervous"), valence (rated on a 1-9 scale from 1: "comfortable" to 9: "uncomfortable"), and fear (rated on a 1-9 scale from 1: "not afraid" to 9: "very afraid") related to the visual stimuli. The questionnaires also included questions about US perception and CS-US contingency. Additionally, participants completed standardized questionnaires before the start of the experiment to assess handedness, depression, anxiety, and stress levels. All participants were screened for contraindications of 7T MRI.

Skin conductance analysis

To eliminate high frequency artefacts, skin conductance data was first low-pass filtered at 10 Hz (62^{nd}-order Blackman FIR filter) in MATLAB (Release 2022b, *RRID:SCR_001622*, The MathWorks Inc., Natick, MA). SCRs were defined as the maximum trough-to peak amplitude within a given time interval after CS onset using semi-automated peak detection implemented in a MATLAB-based EDA-Analysis App[85]. A minimum amplitude criterion of 0.01 µS was used as the SCR detection threshold. In each trial, SCRs were evaluated for two distinct time windows: the conditioned response within a time window of 1.0 s to 5.9 s after CS onset and the unconditioned response within a time window of 6 s to 10 s after CS onset (irrespective of whether a US was presented). To account for between-subject variance, the resulting raw SCR amplitudes were increased by 1 µS and normalized through a logarithmic transformation (LN(1+SCR))[86].

Pupil size analysis

Preprocessing of the raw pupil size data was performed to detect and remove blinks[87]. Trials with fewer than 50% of their data points remaining after blink removal were excluded from the analysis. For each trial, the baseline was computed as the mean pupil size during the 500 ms period prior to CS onset. The baseline was subtracted from the corresponding pupil size[88,89] and the result was divided by the baseline to compute the pupil size response. The mean pupil size response for CS+ and CS- trials were calculated in the time windows from 4 to 5.9 s and 6.0 to 8.0 s for the CS and US pupil size response, respectively. Pupil size responses during possible US

presentation were excluded (5.9 to 6 s). The time window for the CS was chosen based on Jentsch et al.[88] who observed the largest differentiation between CS+ and CS- 2 s prior to the US. The US time window was chosen to match the CS time window. For valid trials, a missing pupillometric data point was considered as 'not a number value' to not influence the pupil size mean computation[89].

Behavioral statistical analysis

For physiological data, due to non-normal distribution (SCR and PSR: Shapiro-Wilk test, $p < 0.001$), statistical analysis was performed using a non-parametric ANOVA-type statistic (ATS) implemented via the ANOVAF option in the PROC Mixed procedure in SAS Studio 3.8 (SAS Institute Inc., Cary, NC, USA). This method is suggested for addressing skewed distributions, outliers, or small sample sizes. To ensure more reliable results, an ATS was applied, with the denominator degrees of freedom set to infinity[90–92] as the use of finite denominator degrees of freedom can result in increased type I errors[93]. The ATS was calculated for each phase (habituation, acquisition training, extinction training, recall test, reacquisition, reextinction, unexpected US phase), with stimulus type (CS+, CS-) and block (early vs. late, defined as the first and second halves of the phase, respectively) as within-subjects factors and SCRs or PSRs as dependent variables.

For analyses of self-reports, an ATS with repeated measures was calculated with stimulus type and phase as within-subjects factors and ratings (valence, arousal, fear and US expectancy) as dependent variables.

For behavioral results, a $p < 0.05$ criterion was used to determine if results were significantly different from those expected if the null hypotheses were correct. Post hoc comparisons were calculated using least square means tests and were adjusted for multiple comparisons using the Tukey-Kramer method.

Computational modeling

An artificial agent was trained to predict the likelihood of a US for a given visual input in a virtual version of the experiment as in Batsikadze et al.[30]. The model was based on reinforcement learning[94] and consisted of a deep neural network (DNN)[95]. The model hyper-parameters were fit to SCRs recorded in the experiment, which served as a proxy for participants' expectation of an US.

The same simplified visual stimuli for trial stimuli s_t and the same encoding for reinforcement signals r_t were used as in Batsikadze et al.[30]. We employed the same network architecture, which comprised two hidden fully connected layers with 64 units each and an output layer with activation function φ. To account for the counterbalanced order of CS+ and CS-presentations in the experimental design, two trial sequences were modeled separately: In trial sequence 1, habituation, extinction, and recall began with a CS+ trial, whereas in trial sequence 2 these phases began with a CS-- trial. For each of the two trial sequences experienced by the participants, 25 randomly initialized agents were trained at each point of the hyper-parameter grid search. Their outputs were averaged and compared to the group-averaged SCR learning curves, and the final set of hyper-parameters were selected by minimizing a cost function that quantified the difference between model predictions and SCR data. After selecting the best-fitting hyper-parameters, 100 randomly initialized agents were trained anew, and their averaged outputs provided the trial-by-trial prediction v_t and prediction errors δ_t used in the fMRI analyses.

On each trial t, the agent stored an experience tuple $e_t = (s_t, r_t, \delta_t)$ in memory for later replay[96]. The agents were trained using the backpropagation algorithm[97] on batches of experiences of size b, which were sampled randomly from memory with a probability that was proportional to a priority score p:

$$P(e_k) = \frac{p_k}{\sum_{j=1}^{T} p_j}$$

Priority scores depended on the experiences' recency λ^{τ}, where τ is the time passed since the experience and λ is a decay factor. Optionally, the priority could additionally depend on the magnitude of US prediction error, i.e., $p = |\delta| \lambda^{\tau}$ [98]. The parameter RPE ∈ {Yes, No} indicates

whether the replay priorities also depended on the magnitude of the US prediction error. The number of replays i was varied to control the degree of learning on each trial.

While the previous model could account for ABA renewal[30], it cannot account for spontaneous recovery in the AAA paradigm, because, in the model, extinction in the same context overwrites the association formed during acquisition. Hence, we extended the model by an additional replay phase, which takes place between habituation, acquisition, extinction, recall and reacquisition phases, and serves to recover the initially acquired association. This replay phase prioritized experiences according to the reinforcement r_k received in a trial and consisted of a total 100 replays of batches of size 128. Reactivation probabilities for these replays were computed as follows:

$$P(e_k) = \frac{e^{\beta r_k}}{\sum_{j=1}^{T} e^{\beta r_j}}$$

where β denotes the inverse temperature and controls the relative difference of reactivation probabilities.

Hyper-parameter fitting

We averaged SCRs from CS+ and CS- trials separately for each trial sequence and applied min-max scaling. The averaged SCRs were accordingly defined as $\bar{Y}_l = (\bar{y}_{+,1}, \ldots, \bar{y}_{+,N}, \bar{y}_{-,1}, \ldots, \bar{y}_{-,N})$, 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 l, respectively. Analogously, the averaged US predictions of the model were defined as $\bar{V}_l(b, \lambda, i, \beta, RPE, \varphi) = (\bar{v}_{+,1}, \ldots, \bar{v}_{+,N}, \bar{v}_{-,1}, \ldots, \bar{v}_{-,N})$, where $\bar{v}_{+,n}$ and $\bar{v}_{-,n}$ are the averaged US predictions for the n-th CS+ and CS- presentations across all model instances who were trained on a given trial sequence l, respectively. For each trial sequence, 25 randomly initialized agents were trained with the same hyper-parameters, and their outputs were averaged to reduce variability and obtain stable trial-wise estimates of US predictions which are able to show a good fit with the population data and prediction errors. The goodness of fit was defined as:

$$\bar{F}(b, \lambda, i, \beta, RPE, \varphi) = \sum_{l=1}^{L} w_l \left(\left\| \bar{Y}_l - \bar{V}_l(b, \lambda, i, \beta, RPE, \varphi) \right\| + P_{Acq} + P_{Gen} + P_{Ext} + P_{Ren} + P_{Rex} \right)$$

where w_l is the number of participants who experienced trial sequence l. 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

1. learn that CS+ is followed by the US: $P_{Acq} = \begin{cases} 0, \bar{v}_{Acq,End} \geq 0.5 \\ 1, \bar{v}_{Acq,End} < 0.5 \end{cases}$ where $\bar{v}_{Acq,End}$ is the average US prediction over the last 2 CS+ presentations of acquisition training.
2. show the CR at the start of extinction training: $P_{Gen} = \begin{cases} 0, \bar{v}_{Ext,Start} \geq 0.5 \\ 1, \bar{v}_{Ext,Start} < 0.5 \end{cases}$ where $\bar{v}_{Ext,Start}$ is the average US prediction over the first 2 CS+ presentations of extinction training.
3. successfully extinguish the CR during extinction training: $P_{Ext} = \begin{cases} 0, \bar{v}_{Ext,End} < 0.1 \\ 1, \bar{v}_{Ext,End} \geq 0.1 \end{cases}$ where $\bar{v}_{Ext,End}$ is the average US prediction over the last 2 CS+ presentations of extinction training.
4. show renewal of the CR: $P_{Ren} = \begin{cases} 0, \bar{v}_{Rec,Start} \geq 0.5 \\ 1, \bar{v}_{Rec,Start} < 0.5 \end{cases}$ where $\bar{v}_{Rec,Start}$ is the average US prediction over the first 2 CS+ presentations of recall.
5. successfully extinguish the CR during recall: $P_{Rex} = \begin{cases} 0, \bar{v}_{Rec,End} \leq 0.1 \\ 1, \bar{v}_{Rec,End} > 0.1 \end{cases}$ where $\bar{v}_{Rec,Start}$ is the average US prediction over the last 2 CS+ presentations of recall

A grid search was conducted over the hyper-parameter sets shown in Table 1. The model with the best goodness-of-fit was then chosen to derive trial-by-trial values for US predictions

and prediction errors. Prediction errors were used for parametric modulation of fMRI data to test hypotheses related to prediction error processing of the cerebellum and VTA.

MRI acquisition

All MRI imaging data were collected while participants were lying supine in the 7T MRI scanner (MAGNETOM Terra, Siemens Healthineers AG, Forchheim, Germany). A 1-channel transmit/32-channel receive head RF coil (Nova Medical Inc., Wilmington MA, USA), was used.

Two dielectric pads for signal homogenization were placed on either side of each participant's upper neck[99]. Depending on head size, further cushioning was added to prevent participant head movement and discomfort.

Whole-brain functional MRI acquisition was performed with a 3-dimensional echo planar image sequence[100] with an isotropic voxel size of 1.5 mm. The sequence was run for each phase (habituation, 98 volumes; acquisition training, 354 volumes; extinction training, 350 volumes; recall test, 182 volumes; reacquisition, reextinction and unexpected US phases; 938 volumes). Imaging parameters were as follows: TR/TE, 1620/20 ms; flip angle, 11°; phase encoding acceleration factor, 2; 3D acceleration factor, 3; slice partial Fourier factor, 7/8; acquisition matrix, 140 × 140; number of slices, 96.

After acquisition training on day 1, a transversal QSM ASPIRE sequence was acquired[101,102] with an isotropic voxel size of 0.7 mm. Further imaging parameters were as follows: TR/TE1/TE2/TE3/TE4/TE5, 28/5/10/15/20/25 ms; flip angle 15°; phase encoding acceleration factor, 3; slice partial Fourier factor, 6/8; acquisition matrix, 320 × 280; number of slices, 256; TA, 10:52 min.

After the acquisition of the QSM ASPIRE sequence, a sagittal MP2RAGE sequence including fat navigators[103–105] was acquired with an isotropic voxel size of 0.75 mm. Further imaging parameters were as follows: TR/TE, 6000/1.85 ms; TI1/TI2, 800/2750 ms; flip angles 1/2, 4°/5°; phase encoding acceleration factor, 3; acquisition matrix, 340 × 340; number of slices, 256; TA, 14:58 min.

Image processing

Motion correction of MP2RAGE volumes including fat navigators was performed using offline reconstruction with the Retro-MoCo toolbox for MATLAB provided by David Gallichan (version23 0.9.0dev, *https://github.com/dgallichan/retroMoCoBox.git*). A T1 map was generated from the motion corrected MP2RAGE with the MP2RAGE-utils package implemented in MATLAB (release 1.0, *https://github.com/srikash/MP2RAGE-utils*). The MP2RAGE was normalized to MNI-space using the CAT12 (CAT12, release 1450)[106] toolbox in SPM12 (Wellcome Department of Cognitive Neurology, London, UK).

Functional MRI volumes were brain extracted with BET (Brain Extraction Tool)[107] in FSL (Release 6.0.1, *RRID:SCR_002823*, Centre for Functional MRI of the Brain, Oxford, UK). Motion and distortion correction was performed for each fMRI run with ANTs (Version 2.3.5, *RRID:SCR_004757*, University of Pennsylvania, Philadelphia, USA)[108]. A 5-volume fMRI sequence preceding each run with phase encoding in the opposite direction was used for the distortion correction. All volumes were coregistered to a T1 map derived from the acquired MP2RAGE with ANTs. The CAT12 processing of the MP2RAGE was used to normalize fMRI data. A 4.5 mm Gaussian kernel was used for smoothing. Motion nuisance regressors (3 translations, 3 rotations) were derived from ANTS affine transformation matrices output.

fMRI analysis

The 1st level analysis was done in MNI space, defining conditions for both the CS and US window (i.e., CS+, CS-, US post CS+, no US post CS+, no US post CS-) separately for each phase. Regressors were calculated from the conditions in SPM12 by convolving a canonical hemodynamic response function with events, which were modelled as delta functions in an event-based design (i.e., the duration of each event is 0 s). Specifically, CS events were time-locked to CS onset (appearance of the visual stimulus), whereas US related events were time-locked to CS offset (6 s after CS onset), which coincided with either delivery or omission of the US. This event-related separation allowed us to test CS related contrasts at CS onset, including CS differentiation (CS+ > CS-). US-related

Grid Search Parameters	
b (batch size)	{1, 16, 32, 64}
λ (decay factor)	{0.5, 0.525, 0.55, ..., 0.95, 0.975, 1}
i (training repeats)	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
β (inverse temperature)	{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5}
RPE	{Yes, No}
φ	{ELU, Sigmoid}

Table 1. Modeling.

Parameters the grid search was run for.

contrasts were tested at CS offset and included contrasts reflecting unexpected omission of the US (e.g. first 3 no US post CS+ > no US post CS-) and US presentation (US post CS+ > no US post CS-). Beta weights were fitted for each regressor by a general linear model for each voxel fMRI time series. Contrasts were computed as linear combinations of beta weights. Values in contrast maps were converted to t-statistic values in first level analysis. Contrast maps are combined and used in a one-sample t-test for second level analysis in SPM12. Tests for significance ($p < 0.05$) were done after threshold-free cluster enhancement (TFCE toolbox in SPM12, R174, *http://dbm.neuro.uni-jena.de/tfce/*) and family-wise error (FWE) corrections. Any trend-level results refer to significance tests without TFCE and FWE corrections ($p < 0.05$). For the cerebellar cortex, results were visualized on cerebellar flatmaps using the SUIT toolbox in SPM12 [22]. For the cerebellar nuclei and VTA, results were visualized in coronal slices and 3D renderings generated in MRIcroGL [109] using custom probabilistic atlases (described in "Volume of interest definition"). To acquire cerebellar anatomical region labels, activation maps were projected onto the SUIT atlas volume (Cerebellum-SUIT.nii, Diedrichsen, 2006).

Event-based analysis

In addition to main effects of conditions (e.g., during acquisition training), events were additionally separated over time for each experimental phase. For fear acquisition training, extinction training, recall test, reacquisition, reextinction and the unexpected US phase, events were grouped in two equal-size blocks representing early and late halves of those phases (e.g., the 8 first CS+ trials of fear acquisition training correspond to early acquisition and the 8 last CS+ trials of fear acquisition training correspond to late acquisition), with equal numbers of CS+ and CS− trials in each block. Trials with unexpected events—such as unexpected US presentations or omissions—were treated as separate single trial regressors. These included the initial three trials of acquisition training, extinction training, recall test, reacquisition and reextinction and specific events such as the three unexpected US omissions in reacquisition. When CS+ trials were modeled as single trials in the GLM, CS- trials were modeled in the same way to maintain a consistent model structure. This modeling choice did not translate into paired or numerically matched CS+ and CS− trials being used in the event-based contrasts. Rather, event-based contrasts for unexpected US omissions contrasted the first three unexpected omissions following CS+ against all "expected" omissions following CS-(e.g., [first 3 no US post CS+ > no US post CS-]).

Parametric modulation analysis

Parametric modulation of fMRI data was performed with prediction and prediction error values in all learning phases derived from our computational model. CS+ and CS- events were modulated with prediction values, while US omission and presentation events were modulated with prediction error values (Figure 9). Events were not further divided within phases (e.g., no early or late halves or separation of the first three trials). Modulations were done separately for each phase, except for reacquisition, reextinction and the unexpected US phase which were modulated together as they were part of the same fMRI run.

Volumes of interest (VOI) definition

A global conjunction between unexpected US omission parametric modulation contrasts during extinction, the recall test and reacquisition was multiplied with a cerebellum mask and showed a region in lobule VI and Crus I. This region was used as a cerebellar cortex (lobule VI and Crus I; Figure S4) volume of interest (VOI). Deep cerebellar nuclei (DCN) and VTA VOIs were drawn for each participant. Drawing of the DCN (i.e., the left and right dentate, globose, emboliform and fastigial nuclei) was done by an expert annotator using both MP2RAGE and QSM information in ITK-SNAP (version 3.8.0) [111]. The DCN were combined into a single bihemispheric VOI. Drawing of the VTA was done by adjusting estimated masks with both MP2RAGE and QSM information, using the red nuclei, substantia nigra, the cerebral aqueduct and interpeduncular fossa as landmarks. Initial masks were estimated for the VTA by inversely transforming the probabilistic atlas by Trutti et al. [79] thresholded at a probability of $p > 0.4$. All drawn masks were transformed into MNI space. After transformation to MNI space, a probabilistic atlas was generated by averaging binary masks for each structure across participants. The probabilistic atlas was thresholded so that the volume

of each structure matched the average volume across participants. For drawing of globose, emboliform and fastigial nuclei, the thresholded atlas was inversely transformed from MNI to T1 space in order to add missing nuclei for each participant with lower quality QSM data. The VTA VOI was used as a seed region in the PPI analysis. In addition, the cerebellar cortex (CB), DCN and VTA VOIs were used as nodes in the 2-node dynamic causal modelling (DCM) networks. The DCN and VTA VOIs were used as masks for displaying fMRI activations in coronal slices and 3D renderings.

Functional connectivity: PPI

For the PPI analysis, the VTA VOI was used to extract time series with SPM12 in MNI space. Time series were used as physiological regressors, and psychophysiological regressors were added for each condition using the gPPI toolbox[112]. Both first- and second-level analyses were performed with the same contrasts as defined in the parametric modulation (e.g., no US x prediction error during extinction) analysis. To summarize connectivity in the parametric modulation analysis, all contrasts related to unexpected US omissions were combined in a single contrast within the gPPI toolbox.

Functional connectivity: DCM

To calculate dynamic causal models (DCMs), the cerebellar cortex (CB) and VTA VOIs were used to extract time series in a concatenated version of the event-based and parametric modulation models using SPM12[113–115]. Each condition was defined as input to the CB and VTA nodes (i.e., CS+, CS-, US post CS+, no US post CS+, no US post CS-), while each prediction or prediction error related event-based contrast (i.e., CS+ > CS-, first 3 no US post CS+ > no US post CS-) or parametric modulation (i.e., no US prediction error) could modulate the reciprocal intrinsic connections. Recurrent connections were active but were not modulated. Parametric Empirical Bayes (PEB) was used for analysis on the second level[116]. All DCM and PEB analyses were done within SPM12.

Supplementary information

Skin conductance responses

Non-parametric ANOVA SCR results

Factor	Numerator Df	*F*	*p*
	Habituation		
Stimulus	1	1.19	0.276
Time	1	26.49	**<0.001*** **
Stimulus x Time	1	0.06	0.802
	Fear acquisition training		
Stimulus	1	20.79	**<0.001*** **
Time	1	28.75	**<0.001*** **
Stimulus x Time	1	3.80	0.051
	Extinction training		
Stimulus	1	7.71	**0.006** **
Time	1	25.05	**<0.001*** **
Stimulus x Time	1	3.74	0.053
	Recall		
Stimulus	1	4.99	**0.026***
Time	1	36.74	**<0.001*** **
Stimulus x Time	1	5.53	**0.019***
	Reacquisition		
Stimulus	1	27.20	**<0.001*** **
Time	1	35.44	**<0.001*** **
Stimulus x Time	1	6.80	**0.009** **
	Reextinction		
Stimulus	1	4.47	**0.035***
Time	1	0.08	0.780
Stimulus x Time	1	0.95	0.329
	Unexpected US phase		
Stimulus	1	5.43	**0.020***
Time	1	0.29	0.593
Stimulus x Time	1	2.34	0.126

Table S1. Non-parametric ANOVA-type statistics for skin conductance responses (SCRs). Results are shown separately for habituation, fear acquisition training, extinction training, recall, reacquisition, reextinction, and the unexpected US phase. Factors included Stimulus (CS+ vs. CS-), Time (early vs. late halves of each phase), and the Stimulus x Time interaction. Reported statistics include numerator degrees of freedom, F-values, and p-values. Significance levels are indicated as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

SCRs for first-three and last-three trial analysis

To assess the robustness of behavioral results when restricting analyses to a small number of trials, we additionally re-binned trials into blocks comprising the first three and last three trials of each phase, instead of halves of each phase. Overall, the results were largely consistent with the primary analysis. The corresponding statistical results are summarized in Table S2.

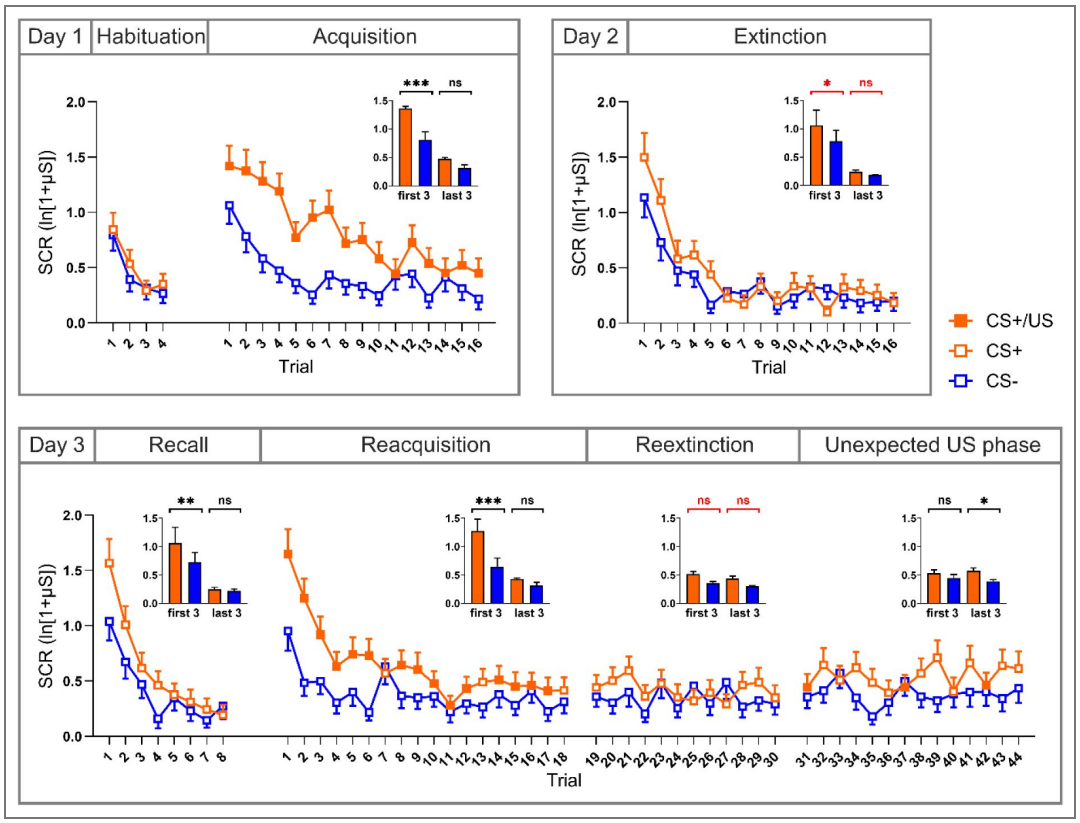

Figure S1. Skin conductance responses (SCRs) for each trial, with CS+ (shown in orange) and CS- (shown in blue) responses paired in blocks. Reinforcement of the CS+ by a US (CS+/US) is indicated by filled squares. The CS- is never reinforced. Bar plots on the top right show mean responses for the first three and last three trials for both CS+ and CS-. On day 1, there was no differentiation between CS+ and CS- in the habituation phase, with significant differentiation emerging during acquisition training. On day 2, CS+/CS- differentiation was confined to the first three extinction trials and was no longer present in the last three extinction trials (trend-level effect; non-significant Stimulus x Time interaction). On day 3, during the initial recall test, participants exhibited spontaneous recovery, i.e., a return of differential responses after extinction training. During initial reacquisition, there were again differential responses to the CS+ and CS-, which decreased in reextinction and the unexpected US phase. Mean values are shown with error bars representing the standard error of the mean. CS: conditioned stimulus; US: unconditioned stimulus; SCR: skin conductance response; µS: microsiemens.

Non-parametric ANOVA SCR results (first-three and last-three trial analysis)

Factor	Numerator Df	*F*	*p*
		Habituation	
Stimulus	1	1.19	0.276
Time	1	26.49	**<0.001*** **
Stimulus x Time	1	0.06	0.802
		Fear acquisition training	
Stimulus	1	8.22	**0.004** **
Time	1	40.06	**<0.001*** **
Stimulus x Time	1	4.05	**0.044***
		Extinction training	
Stimulus	1	4.99	**0.026***
Time	1	34.62	**<0.001*** **
Stimulus x Time	1	3.36	0.067
		Recall	
Stimulus	1	4.24	**0.039***
Time	1	47.42	**<0.001*** **
Stimulus x Time	1	6.35	**0.012***
		Reacquisition	
Stimulus	1	20.47	**<0.001*** **
Time	1	50.62	**<0.001*** **
Stimulus x Time	1	8.70	**0.003** **
		Reextinction	
Stimulus	1	9.36	**0.002** **
Time	1	0.02	0.895
Stimulus x Time	1	0.47	0.494
		Unexpected US phase	
Stimulus	1	4.02	**0.045***
Time	1	0.83	0.363
Stimulus x Time	1	4.67	**0.031***

Table S2. Non-parametric ANOVA-type statistics for skin conductance responses (SCRs) based on the first-three and last-three trial analysis. Results are shown separately for habituation, fear acquisition training, extinction training, recall, reacquisition, reextinction, and the unexpected US phase. Factors included Stimulus (CS+ vs. CS-), Time (first three vs. last three trials), and the Stimulus x Time interaction. Significance levels are indicated as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Pupil size responses

Non-parametric ANOVA PSR results

Factor	Numerator Df	*F*	*p*
		Habituation	
Stimulus	1	1.65	0.200
Time	1	0.26	0.609
Stimulus x Time	1	0.04	0.849
		Fear acquisition training	
Stimulus	1	39.51	**<0.001*** **
Time	1	24.31	**<0.001*** **
Stimulus x Time	1	0.58	0.446
		Extinction training	
Stimulus	1	1.12	0.290
Time	1	18.21	**<0.001*** **
Stimulus x Time	1	4.28	**0.038***
		Recall	
Stimulus	1	6.59	**0.010***
Time	1	6.41	**0.011***
Stimulus x Time	1	4.87	**0.027***
		Reacquisition	
Stimulus	1	47.87	**<0.001*** **
Time	1	4.54	**0.033***
Stimulus x Time	1	7.70	**0.006***
		Reextinction	
Stimulus	1	9.10	**0.003***
Time	1	0.83	0.364
Stimulus x Time	1	4.43	**0.035***
		Unexpected US phase	
Stimulus	1	17.35	**<0.001*** **
Time	1	0.83	0.363
Stimulus x Time	1	0.03	0.872

Table S3. Non-parametric ANOVA-type statistics for pupil size responses (PSRs). Results are shown separately for habituation, fear acquisition training, extinction training, recall, reacquisition, reextinction, and the unexpected US phase. Factors included Stimulus (CS+ vs. CS-), Time (early vs. late halves of each phase), and the Stimulus x Time interaction. Significance levels are indicated as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

PSRs for first-three and last-three trial analysis

To assess the robustness of behavioral results when restricting analyses to a small number of trials, we additionally re-binned trials into blocks comprising the first three and last three trials of each phase, instead of halves of each phase. Overall, the results were largely consistent with the primary analysis. The corresponding statistical results are summarized in Table S4.

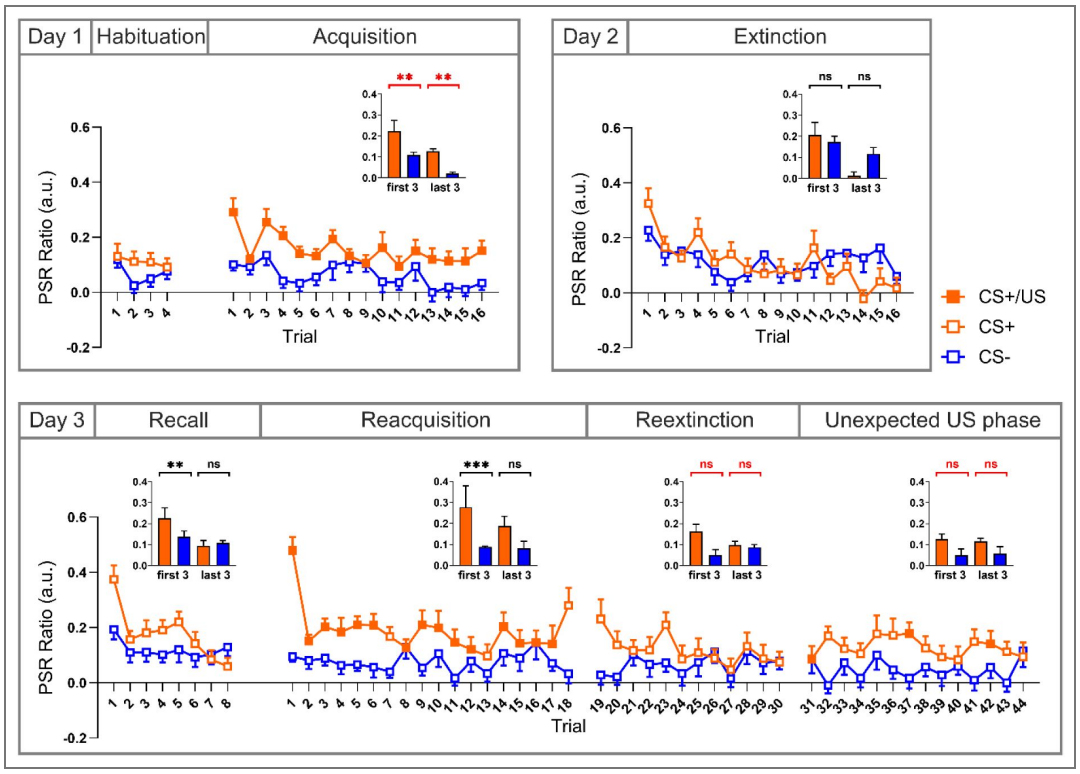

Figure S2. Pupil size responses (PSRs) for each trial, with CS+ (shown in orange) and CS- (shown in blue) responses paired in blocks. Reinforcement of the CS+ by a US (CS+/US) is indicated by filled squares. The CS- is never reinforced. Bar plots on the top right show mean responses for the first three and last three trials for both CS+ and CS-. Significance markers indicate post hoc comparisons between CS+ and CS- within the first three or last three trials; markers are shown in black when the Stimulus x Time interaction was significant and in red when the interaction was not significant. On day 1, there was no differentiation between CS+ and CS- in the habituation phase, with significant differentiation emerging during acquisition training. On day 2, CS+/CS- differentiation decreased across extinction training. Although post hoc CS+/CS- comparisons were not significant during extinction in PSRs, a significant Stimulus x Time interaction was observed, reflecting a decrease from the first three to the last three extinction trials for the CS+ but not for the CS-. On day 3, during initial recall, participants exhibited spontaneous recovery, i.e., a return of differential responses after extinction training. During initial reacquisition, there were again differential responses to the CS+ and CS-, which decreased in reextinction and the unexpected US phase. Mean values are shown with error bars representing the standard error of the mean. CS: conditioned stimulus; US: unconditioned stimulus; PSR: pupil size response.

Non-parametric ANOVA PSR results (first-three and last-three trial analysis)

Factor	Numerator Df	*F*	*p*
		Habituation	
Stimulus	1	1.65	0.200
Time	1	0.26	0.609
Stimulus x Time	1	0.04	0.849
		Fear acquisition training	
Stimulus	1	29.02	**<0.001*** **
Time	1	30.93	**<0.001*** **
Stimulus x Time	1	0.35	0.554
		Extinction training	
Stimulus	1	0.93	0.334
Time	1	41.57	**<0.001*** **
Stimulus x Time	1	7.72	**0.006** **
		Recall	
Stimulus	1	4.72	**0.030***
Time	1	12.23	**<0.001*** **
Stimulus x Time	1	7.13	**0.008** **
		Reacquisition	
Stimulus	1	39.99	**<0.001*** **
Time	1	9.40	**0.002** **
Stimulus x Time	1	8.22	**0.004** **
		Reextinction	
Stimulus	1	4.44	**0.035***
Time	1	0.36	0.547
Stimulus x Time	1	2.63	0.105
		Unexpected US phase	
Stimulus	1	4.36	**0.037***
Time	1	0.16	0.689
Stimulus x Time	1	0.14	0.705

Table S4. Non-parametric ANOVA-type statistics for pupil size responses (PSRs) based on the first three and last three trial analysis. Results are shown separately for habituation, fear acquisition training, extinction training, recall, reacquisition, reextinction, and the unexpected US phase. Factors included Stimulus (CS+ vs. CS-), Time (first three vs. last three trials), and the Stimulus x Time interaction. Significance levels are indicated as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Self-reports

Self-reports results summary

Stimulus	Time of assessment				
	Post habituation	Post acquisition	Post extinction	Post recall	End of day 3
Arousal ratings (1 - very calm, 9 - very excited)					
CS+	2 (1-3)	**6 (5-7)**	**3 (1-6)**	**3 (1-6)**	**7 (4-7)**
CS-	2 (1-4)	**2 (1-3)**	**1 (1-3)**	**1 (1-3)**	**1 (1-2)**
Fear ratings (1 - not afraid, 9 - very afraid)					
CS+	1 (1-2)	**6 (3-6)**	**2 (1-5)**	**2 (1-4)**	**5 (3-7)**
CS-	1 (1-2)	**1 (1-2)**	**1 (1-2)**	**1 (1-2)**	**1 (1-2)**
US expectancy (1 - US not expected, 5 - unsure, 9 - US expected)					
CS+	1 (1-2)	**9 (9-9)**	**3 (1-5)**	**3 (1-5)**	**6 (5-7)**
CS-	1 (1-2)	**1 (1-2)**	**1 (1-1)**	**1 (1-2)**	**1 (1-2)**
Valence ratings (1 - unpleasant, 9 - very pleasant)					
CS+	7 (5-8)	**3 (2-4)**	**5 (4-7)**	**6 (5-8)**	**3 (3-5)**
CS-	7 (5-9)	**8 (6-9)**	**7 (5-9)**	**8 (6-9)**	**8 (6-9)**

Table S5. Summary of self-report ratings. Median values and interquartile ranges (in parentheses) are shown for arousal, fear, US expectancy, and valence ratings for CS+ and CS- assessed after habituation, acquisition training, extinction training, recall test, and at the end of day 3. Rating scales ranged from 1 to 9, with anchors as indicated for each measure. Interquartile ranges are shown in parentheses. Statistically significant differences between CS+ and CS- are indicated in bold (least squares means tests; $p < 0.01$).

Non-parametric ANOVA results for self-reports

Factor	Df	F	p
		Arousal	
Stimulus	1	128.71	**<0.001*****
Time	3.51	16.31	**<0.001*****
Stimulus x Time	3.29	35.06	**<0.001*****
		Fear	
Stimulus	1	105.13	**<0.001*****
Time	3.32	18.60	**<0.001*****
Stimulus x Time	3.4	34.47	**<0.001*****
		Valence	
Stimulus	1	124.79	**<0.001*****
Time	3.39	19.80	**<0.001*****
Stimulus x Time	3.11	51.42	**<0.001*****
		US expectancy	
Stimulus	1	451.62	**<0.001*****
Time	2.4	62.78	**<0.001*****
Stimulus x Time	2.8	187.33	**<0.001*****

Table S6. Non-parametric ANOVA-type statistics for self-report measures. Results are shown for arousal, fear, valence, and US expectancy ratings, with Stimulus (CS+ vs. CS-) and Time of assessment as within-subject factors, as well as the Stimulus x Time interaction. Degrees of freedom, F values, and p values are reported for each effect. Significance levels are indicated as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

fMRI results

fMRI PPI results: VTA connectivity with a cerebellar (CB) seed

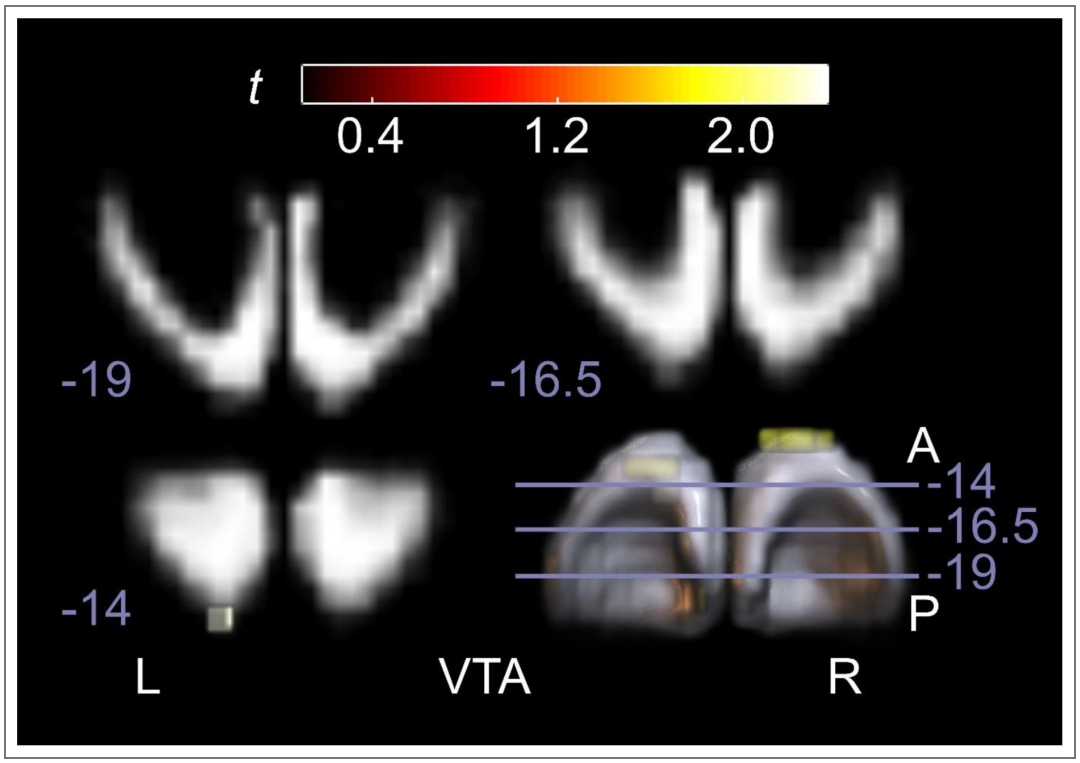

Figure S3. Trend-level psychophysiological interaction (PPI) results using a cerebellar cortex seed (CB; VOI defined from the conjunction analysis; see VOI definition in Methods) during unexpected US omission contrasts. Cerebellar activations are displayed on cerebellar flatmaps (SUIT), while midbrain activations are shown on coronal slices progressing from posterior to anterior, with MNI y-coordinates indicated for each slice. The color scale corresponds to uncorrected voxelwise t-values ($p < 0.05$, uncorrected). Effects in the ventral tegmental area (VTA) are weak and spatially limited, reflecting the reduced sensitivity of cerebellar-seed PPI analyses in the present dataset. This analysis is included for completeness and illustrates connectivity patterns when reversing the seed direction relative to the main VTA-seed-PPI analyses shown in the main figures.

Summary statistics for VOI contrast estimates across fMRI contrasts

Contrast	ROI	Mean	95% CI	Cohen's d
Figure 4: Acquisition, US prediction and presentation				
A: CS+ > CS-	CB	0.173	[-0.373, 0.720]	0.097
	DCN	-0.200	[-0.773, 0.373]	-0.107
	VTA	0.888	**[0.458, 1.319]**	**0.635**
B: CS+ x P (inv)	CB	1.872	**[0.894, 2.849]**	**0.589**
	DCN	0.562	[-0.406, 1.529]	0.179
	VTA	2.078	**[1.281, 2.874]**	**0.803**
C: US presentation	CB	5.367	**[3.879, 6.855]**	**1.110**
	DCN	2.040	**[1.131, 2.950]**	**0.691**
	VTA	2.852	**[2.064, 3.639]**	**1.115**
Figure 5: Extinction, US prediction and omission				
A: CS+ > CS-	CB	-0.057	[-0.566, 0.452]	-0.034
	DCN	-0.143	[-0.649, 0.363]	-0.087
	VTA	0.223	[-0.170, 0.615]	0.175
B: CS+ x Prediction	CB	1.043	[-0.083, 2.169]	0.285
	DCN	0.356	[-0.468, 1.180]	0.133
	VTA	2.057	**[1.248, 2.865]**	**0.783**
C: US presentation	CB	0.471	[-0.062, 1.004]	0.272
	DCN	0.311	[-0.208, 0.829]	0.184
	VTA	0.120	[-0.351, 0.590]	0.078
Figure 6: Unexpected US omission (event-based)				
A: Extinction	CB	4.487	**[2.762, 6.211]**	**0.801**
	DCN	1.989	**[0.653, 3.325]**	0.458
	VTA	1.807	**[0.616, 2.997]**	0.467
B: Recall	CB	4.485	**[2.823, 6.148]**	**0.830**
	DCN	1.352	**[0.072, 2.632]**	0.325
	VTA	1.949	**[0.903, 2.996]**	**0.573**
C: Reacquisition	CB	8.082	**[6.209, 9.955]**	**1.328**
	DCN	0.872	[-0.379, 2.123]	0.215
	VTA	3.711	**[2.133, 5.289]**	**0.724**
D: Reextinction	CB	5.780	**[3.748, 7.811]**	**0.876**
	DCN	0.969	[-0.424, 2.361]	0.214
	VTA	3.152	**[1.699, 4.605]**	**0.668**
Figure 7: Unexpected US omission (parametric modulation)				
A: Extinction	CB	3.409	**[2.034, 4.784]**	**0.763**
	DCN	1.114	**[0.259, 1.970]**	0.401
	VTA	1.871	**[0.966, 2.777]**	**0.636**
B: Recall	CB	4.029	**[2.818, 5.241]**	**1.024**

	DCN	1.360	[-0.217, 2.938]	0.265
	VTA	1.786	**[0.696, 2.877]**	**0.504**
	CB	2.675	**[1.221, 4.128]**	**0.566**
C: Reacquisition	DCN	0.192	[-0.916, 1.301]	0.053
	VTA	1.712	**[0.398, 3.025]**	0.401
	CB	1.505	[-0.742, 3.753]	0.206
D: Reextinction	DCN	0.671	[-1.381, 2.724]	0.101
	VTA	1.256	[-0.551, 3.063]	0.214
Figure 8: Unexpected US omission (PPI, VTA seed)				
A: Extinction	CB	-2.069	[-6.848, 2.710]	-0.133
A: Recall	CB	3.209	[-1.952, 8.370]	0.191
A: Reacquisition	CB	2.558	[-0.613, 5.729]	0.248
A: Reextinction	CB	2.585	[-1.824, 6.994]	0.180
B: All	CB	1.571	[-0.593, 3.735]	0.223

Table S7. Summary statistics for subject-level VOI contrast estimates across fMRI contrasts. For each contrast and region of interest (CB: cerebellar cortex; DCN: deep cerebellar nuclei; VTA: ventral tegmental area), the table reports mean contrast estimates, 95% confidence intervals (CI), and Cohen's d (one-sample effect size relative to zero). Effect sizes with Cohen's d > 0.5 and 95% confidence intervals entirely above zero are highlighted in bold to facilitate interpretation of effect magnitude and consistency across participants. Event-based contrasts show comparatively consistent effects despite being based on a small number of trials (e.g., first three unexpected US omissions), whereas parametric modulation and psychophysiological interaction (PPI) analyses incorporate a larger number of observations and show greater inter-individual variability.

Cerebellar VOI used for connectivity analyses on a SUIT flatmap

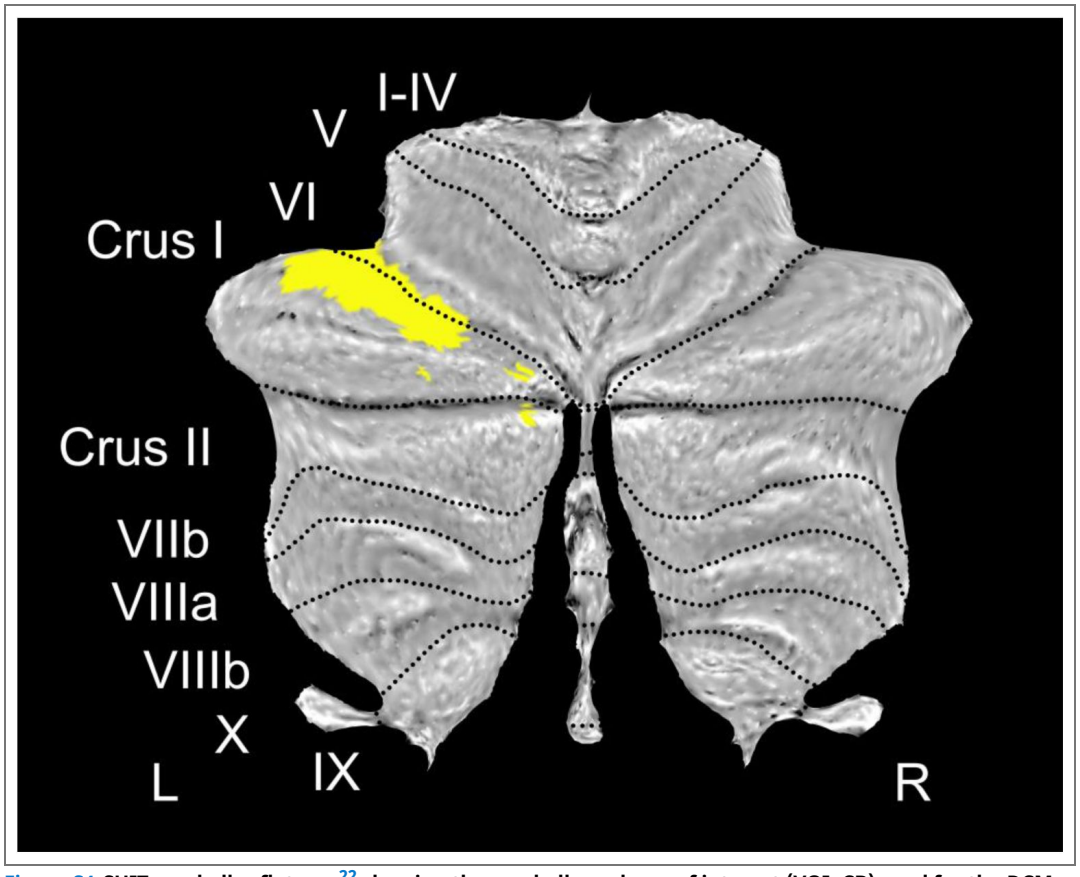

Figure S4. SUIT cerebellar flatmap[22] showing the cerebellar volume of interest (VOI; CB) used for the DCM and PPI analyses. The VOI (yellow) is restricted to the cerebellar cortex and was defined as the conjunction of unexpected US omission contrasts (see Methods: Volumes of interest (VOI) definition). This VOI was used as the

cerebellar node for the DCM (Figure 8) and as the seed region for the PPI analysis (Figure S3).

fMRI activation cluster tables

fMRI activations related to the prediction and presentation of the US. TFCE and FWE corrected

Index	Location (lobule, DCN, VTA)	Side	MNI coordinates/mm			Cluster size (number of voxels)	p_{FWE}	TFCE t
			x	y	z			
Figure 4A: Acquisition, CS+ > CS-, t-test, TFCE, p < 0.05, FWE corr.								
1	VTA	right	4.0	-16.0	-13.0	27	0.016	1599
2	VTA	left	-6.5	-17.5	-11.5	14	0.021	1503
3	VTA	left	-2.0	-17.5	-10.0	1	0.036	1293
Figure 4B: CS+ x prediction (inverted) during acquisition, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left VI (2109), right VI (1758), right V (1324), left Crus I (1315), white matter (802), left V (748), right I-IV (643), left Crus II (506), right Crus I (453), vermal VI (380), left I-IV (366), left VIIb (137), left VIIIa (90), vermal VIIIa (57), vermal IX (53), left IX (46), right DN (40), vermal X (38), right IX (35), vermal VIIIb (33), vermal VIIb (23), right VIIIa (15), right X (7), right VIIb (5), right FN (5), vermal Crus I (3), vermal Crus II (2), left DN (2), left FN (2)						
	VI	right	25.0	-61.0	-16.0	10997	<0.001	6919
	VI	right	31.0	-55.0	-19.0		<0.001	6890
	V	right	19.0	-52.0	-13.0		<0.001	6636
2	Extended cluster	right VTA (70), left VTA (67)						
	VTA	left	-2.0	-16.0	-14.5	137	0.001	3389
	VTA	right	8.5	-19.0	-13.0		0.001	3202
Figure 4C: US post CS+ > no US post CS- during acquisition, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left Crus I (4514), white matter (4220), left VI (3448), right VI (3362), right Crus I (2954), left Crus II (2681), right Crus II (2223), right V (1615), left VIIb (1503), right VIIb (1494), right VIIIa (1382), right I-IV (1368), left V (1251), left I-IV (1147), left VIIIa (1102), right VIIIb (1054), left VIIIb (889), right IX (816), vermal VI (806), left IX (723), vermal VIIIa (452), left DN (308), vermal IX (297), right DN (283), left X (267), right X (252), vermal VIIIb (226), vermal Crus II (168), vermal X (98), vermal VIIb (69), left IN (24), right IN (24), left FN (9), vermal Crus I (8), right FN (8)						
	VI	Right	31.0	-70.0	-22.0	41045	<0.001	16742
	VIIb	Right	20.5	-70.0	-49.0		<0.001	16480
	VI	Right	23.5	-68.5	-28.0		<0.001	16166
2	VTA	Left	-0.5	-16.0	-14.5	156	0.001	13357

	VTA	Right	7.0	-16.0	-10.0		0.001	13090
Figure 5A: CS+ > CS- during extinction, t-test, TFCE, p < 0.05, FWE corr.								
	No significant voxels							
Figure 5B: CS+ x prediction during extinction, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left VTA (72), right VTA (63)						
	VTA	right	7.0	-16.0	-11.5	135	0.002	2603
	VTA	left	-3.5	-16.0	-13.0		0.003	2495
2	Extended cluster	right VI (894), left VI (805), right V (410), left Crus I (197), vermal VI (167), left V (40), right I-IV (17), white matter (14), right Crus I (5), left I-IV (3)						
	VI	left	-27.5	-67.0	-19.0	2552	0.004	2328
	VI	right	31.0	-65.5	-19.0		0.004	2298
	VI	right	31.0	-55.0	-19.0		0.004	2295
3	I-IV	left	-3.5	-46.0	-4.0	10	0.009	1990
4	V	left	-17.0	-44.5	-13.0	1	0.029	1457
5	I-IV	right	4.0	-53.5	-10.0	4	0.05	1223
Figure 5C: No US post CS+ > no US post CS- during extinction, t-test, TFCE, p < 0.05, FWE corr.								
	No significant voxels							

Table S8. fMRI activation clusters related to the prediction, presentation and omission of the unconditioned stimulus (US) during acquisition and extinction training (Figure 4 and 5). Clusters were identified in the cerebellar cortex, deep cerebellar nuclei (DCN), and ventral tegmental area (VTA) using threshold-free cluster enhancement (TFCE) with family-wise error (FWE) correction ($p < 0.05$). Up to three local maxima per cluster are reported, separated by at least 8 mm. Coordinates are given in MNI space (x, y, z). Cluster size is reported as number of voxels (voxel volume = 3.375 mm^3). US: unconditioned stimulus; CS: conditioned stimulus; VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; MNI: Montreal Neurological Institute standard brain; TFCE t: threshold-free cluster-enhanced t-statistic; pFWE: family-wise error-corrected p-value.

fMRI activations related to the prediction and presentation of the US. Uncorrected

Index	Location (lobule, DCN, VTA)	Side	MNI coordinates/mm			Cluster size (number of voxels)	p_{unc}	t
			x	y	z			
Figure 4A: CS+ > CS- during acquisition, t-test, p < 0.05, uncorrected								
1	Extended cluster	right VI (701), right Crus I (259), white matter (4), right V (2), right VIIIa (2), right X (1)						
	VI	right	35.5	-53.5	-29.5	969	<0.001	5.05
	VI	right	34.0	-34.0	-35.5		<0.001	4.13
	VI	right	28.0	-65.5	-22.0		0.001	3.21
2	Extended cluster	left VI (425), left Crus I (386), left Crus II (2)						
	VI	left	-29.0	-62.5	-22.0	813	<0.001	4.48
	Crus I	left	-47.0	-44.5	-32.5		<0.001	3.82
	VI	left	-35.0	-55.0	-28.0		<0.001	3.70

3	Extended cluster	right VTA (62), left VTA (47)						
	VTA	right	4.0	-16.0	-13.0	109	<0.001	4.31
	VTA	left	-6.5	-17.5	-11.5		<0.001	4.30
4	Extended cluster	left Crus I (346), left VI (144), left Crus II (131), vermal VI (96), right VI (51), white matter (6)						
	VI	right	8.5	-76.0	-16.0	774	<0.001	3.97
	Crus I	left	-18.5	-80.5	-31.0		<0.001	3.56
	Crus II	left	-5.0	-77.5	-38.5		0.007	2.57
5	Extended cluster	left VIIb (369), left Crus II (44), left VIIIa (30)						
	VIIb	left	-33.5	-64.0	-58.0	443	0.001	3.40
	VIIb	left	-42.5	-59.5	-55.0		0.001	3.23
	VIIb	left	-38.0	-52.0	-50.5		0.004	2.78
6	Extended cluster	right VIIb (150), right VIIIa (146)						
	VIIIa	right	34.0	-58.0	-55.0	296	0.001	3.39
	VIIb	right	40.0	-50.5	-53.5		0.001	3.17
7	Extended cluster	left IX (117), vermal IX (63), white matter (14), left VIIIb (4), vermal X (4)						
	IX	left	-8.0	-56.5	-40.0	202	0.001	3.18
	IX	vermal	-2.0	-55.0	-32.5		0.008	2.51
8	Extended cluster	right VIIIa (94), right VIIb (46)						
	VIIIa	right	8.5	-71.5	-53.5	140	0.002	2.99
	VIIIa	right	22.0	-65.5	-56.5		0.003	2.87
9	Extended cluster	left VIIIa (44), left VIIb (15)						
	VIIb	left	-18.5	-71.5	-58.0	59	0.003	2.94
	VIIIa	left	-11.0	-68.5	-56.5		0.033	1.87
10	I-IV	left	-0.5	-49.0	-20.5	60	0.005	2.69
11	IX	right	10.0	-58.0	-47.5	75	0.005	2.68
12	V	left	-0.5	-62.5	-1.0	19	0.008	2.52
13	Crus II	right	4.0	-83.5	-40.0	24	0.011	2.38
14	white matter		28.0	-52.0	-46.0	16	0.014	2.28
15	Crus II	right	46.0	-59.5	-50.5	14	0.015	2.23
16	VIIIb	right	20.5	-59.5	-49.0	33	0.018	2.16
17	VIIIb	left	-12.5	-58.0	-49.0	10	0.028	1.97
18	X	left	-20.0	-34.0	-47.5	3	0.032	1.90
19	Crus II	right	10.0	-89.5	-37.0	2	0.033	1.87
20	VIIIa	left	-9.5	-70.0	-55.0	2	0.033	1.87
21	IX	right	8.5	-49.0	-55.0	2	0.04	1.78
22	VIIIb	left	-20.0	-37.0	-50.5	1	0.042	1.76
23	IX	right	10.0	-55.0	-55.0	4	0.042	1.76
24	white matter		10.0	-71.5	-35.5	4	0.043	1.75
25	VIIIa	right	31.0	-46.0	-55.0	1	0.044	1.73
26	white matter		-9.5	-50.5	-28.0	1	0.045	1.73
27	I-IV	left	-2.0	-56.5	0.5	1	0.045	1.73

28	Crus I	right	26.5	-83.5	-22.0	1	0.046	1.71
29	Crus II	left	-41.0	-46.0	-46.0	2	0.046	1.71
30	VIIIa	left	-6.5	-71.5	-52.0	1	0.047	1.71
31	V	left	-29.0	-31.0	-31.0	1	0.047	1.70
Figure 4B: CS+ x prediction during acquisition, t-test, p < 0.05, uncorrected								
1	Extended cluster	left VI (2607), right VI (2099), left Crus I (1857), right V (1449), white matter (1256), left V (944), right I-IV (886), right Crus I (780), left Crus II (651), left I-IV (510), vermal VI (436), left VIIb (182), right Crus II (181), vermal IX (140), left VIIIa (120), vermal VIIIa (111), right VIIb (110), left IX (82), right IX (80), right DN (77), vermal VIIIb (61), vermal X (53), right VIIIa (39), vermal VIIb (38), right VIIIb (27), left DN (19), right X (13), vermal Crus II (12), right FN (7), vermal Crus I (4), left FN (4), left IN (3)						
	VI	right	25.0	-59.5	-16.0	14838	<0.001	6.47
	VI	right	31.0	-55.0	-19.0		<0.001	6.32
	V	right	1.0	-62.5	-1.0		<0.001	6.15
2	Extended cluster	right VTA (72), left VTA (69)						
	VTA	left	-2.0	-20.5	-17.5	141	<0.001	5.05
	VTA	left	-2.0	-16.0	-8.5		<0.001	4.78
	VTA	right	8.5	-19.0	-13.0		0.001	3.45
3	Extended cluster	right Crus II (258), right VIIb (210), right Crus I (14), right VIIIa (11)						
	Crus II	right	43.0	-55.0	-47.5	493	<0.001	3.76
	Crus II	right	43.0	-67.0	-53.5		0.003	2.85
	VIIb	right	31.0	-71.5	-50.5		0.011	2.38
4	Extended cluster	left VIIIa (82), left VIIb (12)						
	VIIIa	left	-27.5	-56.5	-56.5	94	0.002	2.99
	VIIIa	left	-36.5	-53.5	-58.0		0.04	1.78
5	VIIb	left	-30.5	-74.5	-55.0	80	0.003	2.85
6	IX	left	-11.0	-49.0	-46.0	89	0.004	2.75
7	VIIIb	right	23.5	-46.0	-58.0	16	0.007	2.56
8	Extended cluster	left Crus I (73), left Crus II (9)						
	Crus I	left	-45.5	-67.0	-43.0	82	0.008	2.52
	Crus I	left	-36.5	-61.0	-40.0		0.021	2.09
9	IX	right	8.5	-50.5	-59.5	29	0.009	2.44
10	Crus II	right	28.0	-85.0	-47.5	27	0.01	2.41
11	VIIIa	right	20.5	-62.5	-52.0	45	0.012	2.35
12	VIIIb	left	-24.5	-44.5	-58.0	13	0.014	2.28
13	VIIIb	left	-24.5	-38.5	-53.5	8	0.014	2.28
14	I-IV	left	-11.0	-46.0	-7.0	2	0.016	2.21
15	Crus I	right	37.0	-65.5	-41.5	29	0.02	2.12
16	I-IV	left	-0.5	-52.0	2.0	1	0.022	2.07
17	I-IV	left	-0.5	-46.0	-5.5	8	0.028	1.96
18	X	right	23.5	-34.0	-47.5	1	0.036	1.84
19	I-IV	right	8.5	-46.0	-4.0	1	0.037	1.83
20	IX	left	-8.0	-52.0	-59.5	3	0.038	1.81

21	VIIb	left	-41.0	-52.0	-55.0	2	0.039	1.80
22	VIIIa	right	10.0	-70.0	-56.5	1	0.046	1.71
23	Crus I	left	-53.0	-55.0	-35.5	1	0.048	1.69
24	VIIIb	right	16.0	-44.5	-50.5	1	0.049	1.68
Figure 4C: US post CS+ > no US post CS- during acquisition, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (4378), white matter (3746), left VI (3367), right VI (3353), right Crus I (2720), left Crus II (2536), right Crus II (2069), right V (1592), left VIIb (1456), right VIIb (1455), right VIIIa (1352), right I-IV (1314), left V (1157), left I-IV (1098), right VIIIb (1034), left VIIIa (983), left VIIIb (828), vermal VI (799), right IX (748), left IX (668), vermal VIIIa (452), left DN (308), vermal IX (296), right DN (278), left X (257), right X (246), vermal VIIIb (224), vermal Crus II (168), vermal X (83), vermal VIIb (69), left IN (24), right IN (24), left FN (9), vermal Crus I (8), right FN (8)							
	VIIb	right	20.5	-70.0	-49.0	39107	<0.001	10.43
	VI	right	31.0	-70.0	-22.0		<0.001	9.82
	VIIb	left	-17.0	-70.0	-49.0		<0.001	8.89
2	Extended cluster	right VTA (81), left VTA (75)						
3	VTA	left	-0.5	-16.0	-14.5	156	<0.001	6.77
	VTA	right	7.0	-16.0	-10.0		<0.001	5.07
	IX	right	4.0	-46.0	-43.0	5	0.012	2.35
Figure 5A: CS+ > CS- during extinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	right VIIb (96), right VIIIa (53), right Crus II (3)						
	VIIb	right	34.0	-73.0	-56.5	152	<0.001	3.62
	VIIIa	right	31.0	-58.0	-52.0		0.016	2.21
2	VIIb	right	23.5	-73.0	-58.0	16	0.004	2.80
3	Crus II	left	-15.5	-80.5	-46.0	50	0.004	2.76
4	VIIIa	left	-29.0	-49.0	-49.0	49	0.006	2.64
5	VIIb	left	-33.5	-62.5	-53.5	139	0.006	2.60
6	Crus II	right	13.0	-77.5	-46.0	38	0.007	2.56
7	Extended cluster	white matter (25), right Crus II (20), right Crus I (16), right VI (15), right VIIb (14), right VIIIa (1)						
	Crus II	right	38.5	-49.0	-44.5	91	0.008	2.53
	VI	right	32.5	-46.0	-37.0		0.012	2.36
8	Crus I	right	50.5	-61.0	-41.5	12	0.015	2.25
9	VI	left	-33.5	-34.0	-32.5	19	0.016	2.21
10	VIIIa	right	23.5	-62.5	-56.5	14	0.016	2.20
11	white matter		-18.5	-31.0	-37.0	3	0.02	2.11
12	X	right	16.0	-40.0	-47.5	5	0.021	2.10
13	white matter		20.5	-34.0	-40.0	12	0.024	2.04
14	VIIb	right	41.5	-50.5	-53.5	2	0.026	1.99
15	white matter		-12.5	-37.0	-40.0	4	0.027	1.98
16	IX	right	8.5	-53.5	-55.0	13	0.027	1.98
17	VTA	right	5.5	-19.0	-16.0	5	0.027	1.98

18	VIIb	left	-11.0	-74.5	-55.0	2	0.028	1.96
19	VI	right	8.5	-76.0	-16.0	6	0.029	1.94
20	VI	left	-15.5	-77.5	-20.5	7	0.03	1.93
21	VI	left	-29.0	-58.0	-29.5	5	0.03	1.92
22	IX	left	-5.0	-52.0	-44.5	4	0.034	1.87
23	IX	left	-6.5	-59.5	-53.5	2	0.036	1.83
24	white matter		14.5	-40.0	-40.0	1	0.036	1.83
25	VI	right	37.0	-37.0	-34.0	2	0.037	1.82
26	X	left	-27.5	-35.5	-43.0	1	0.038	1.81
27	VTA	right	1.0	-19.0	-10.0	3	0.038	1.81
28	VI	left	-20.0	-67.0	-25.0	6	0.038	1.81
29	VI	left	-36.5	-40.0	-28.0	2	0.039	1.80
30	VIIIb	left	-11.0	-56.5	-56.5	5	0.039	1.80
31	VIIIb	vermal	1.0	-65.5	-49.0	2	0.042	1.76
32	VIIb	left	-11.0	-77.5	-52.0	1	0.042	1.76
33	Crus I	left	-39.5	-50.5	-40.0	1	0.043	1.75
34	VI	right	16.0	-74.5	-16.0	1	0.044	1.73
35	white matter		-11.0	-41.5	-38.5	1	0.047	1.70
36	Crus II	left	-23.0	-79.0	-50.5	1	0.047	1.70

Figure 5B: CS+ x prediction during extinction, t-test, p < 0.05, uncorrected

1	Extended cluster	left VTA (75), right VTA (74)						
	VTA	right	7.0	-16.0	-11.5	149	<0.001	5.64
	VTA	left	-5.0	-16.0	-13.0		<0.001	4.90
2	Extended cluster	left VI (1484), right VI (1377), white matter (711), right V (627), left Crus I (490), right IX (488), left IX (467), right VIIIa (410), right VIIb (380), right VIIIb (246), left VIIIb (234), vermal VI (225), right I-IV (183), right Crus II (178), left I-IV (155), vermal IX (143), vermal X (104), left V (85), right Crus I (84), left X (63), vermal VIIIb (47), right X (46), right DN (37), left DN (32), vermal VIIIa (27), left Crus II (22), left IN (18), left VIIIa (12), left VIIb (7), left FN (6), right FN (3), right IN (2)						
	white matter		5.5	-68.5	-11.5	8393	<0.001	5.03
	VIIIb	left	-18.5	-43.0	-46.0		<0.001	4.68
	VI	left	-5.0	-77.5	-17.5		<0.001	4.61
3	Extended cluster	left Crus II (198), left VIIb (55), vermal Crus II (9)						
	Crus II	left	-0.5	-80.5	-38.5	262	0.002	3.07
	VIIb	left	-12.5	-76.0	-47.5		0.004	2.80
	Crus II	left	-21.5	-76.0	-47.5		0.024	2.03
4	I-IV	left	-3.5	-46.0	-4.0	25	0.002	3.04
5	Extended cluster	left VIIb (162), left VIIIa (51)						
	VIIb	left	-32.0	-61.0	-55.0	213	0.006	2.65
	VIIb	left	-38.0	-55.0	-52.0		0.007	2.58
	VIIb	left	-21.5	-71.5	-56.5		0.01	2.43
6	X	right	20.5	-34.0	-46.0	12	0.007	2.57

7	V	left	-17.0	-44.5	-13.0	5	0.01	2.40
8	Crus I	right	41.5	-53.5	-37.0	13	0.011	2.38
9	Crus I	right	40.0	-74.5	-22.0	13	0.024	2.03
10	X	left	-18.5	-34.0	-47.5	5	0.025	2.01
11	Crus II	left	-32.0	-74.5	-50.5	9	0.025	2.01
12	V	left	-21.5	-46.0	-16.0	1	0.034	1.87
13	DN	right	16.0	-59.5	-38.5	3	0.034	1.87
14	V	left	-23.0	-50.5	-16.0	1	0.038	1.81
15	Crus I	right	47.5	-53.5	-34.0	1	0.038	1.81
16	VI	vermal	5.5	-71.5	-25.0	4	0.038	1.81
17	Crus II	left	-33.5	-77.5	-53.5	2	0.039	1.80
18	Crus II	right	34.0	-76.0	-52.0	3	0.039	1.79
19	white matter		-17.0	-31.0	-35.5	2	0.039	1.79
20	I-IV	right	2.5	-41.5	-16.0	1	0.041	1.77
21	Crus I	left	-45.5	-43.0	-31.0	1	0.044	1.74

Figure 5C: No US post CS+ > no US post CS- during extinction, t-test, p <0.05, uncorrected

1	Extended cluster	left Crus I (1015), white matter (539), left VI (451), left Crus II (386), left VIIb (58), left VIIIa (34), left DN (34), left V (22), left IX (8)						
	VI	left	-24.5	-55.0	-31.0	2547	<0.001	4.20
	Crus I	left	-41.0	-58.0	-37.0		<0.001	3.84
	white matter		-30.5	-52.0	-37.0		<0.001	3.71
2	Extended cluster	right V (191), left V (160), left VI (109), vermal VI (97), right VI (86), right Crus I (40), right I-IV (14), white matter (2)						
	V	right	2.5	-67.0	-5.5	699	<0.001	3.73
	VI	right	8.5	-80.5	-20.5		0.001	3.33
	V	left	-6.5	-62.5	-10.0		0.001	3.30
3	Extended cluster	right V (207), right VI (199), white matter (194), right I-IV (126), right Crus I (12)						
	V	right	28.0	-41.5	-28.0	738	0.001	3.33
	white matter		20.5	-40.0	-31.0		0.001	3.18
	I-IV	right	10.0	-47.5	-25.0		0.002	3.14
4	Extended cluster	white matter (120), right VIIb (115), right Crus II (68), right VIIIa (58), right Crus I (25)						
	white matter		28.0	-49.0	-41.5	386	0.001	3.17
	VIIb	right	38.5	-47.5	-46.0		0.002	3.09
	VIIIa	right	32.5	-52.0	-53.5		0.006	2.65
5	VIIIb	right	23.5	-50.5	-56.5	38	0.002	3.07
6	Extended cluster	right Crus I (311), right VI (222)						
	VI	right	22.0	-68.5	-28.0	533	0.002	2.99
	VI	right	32.5	-67.0	-26.5		0.003	2.84
	VI	right	28.0	-59.5	-28.0		0.012	2.33
7	IX	right	5.5	-49.0	-59.5	11	0.003	2.88

8	Extended cluster	white matter (119), left DN (20), left I-IV (15), left IN (13), left V (12)						
	DN	left	-9.5	-56.5	-29.5	179	0.003	2.87
	I-IV	left	-8.0	-47.5	-25.0		0.016	2.21
9	I-IV	right	14.5	-43.0	-11.5	15	0.005	2.71
10	Crus I	left	-51.5	-67.0	-37.0	15	0.005	2.68
11	VTA	right	2.5	-16.0	-13.0	19	0.008	2.53
12	VIIIa	right	17.5	-65.5	-58.0	27	0.008	2.50
13	Extended cluster	left VIIb (40), left VIIIa (36)						
	VIIb	left	-38.0	-62.5	-59.5	76	0.009	2.48
	VIIIa	left	-32.0	-52.0	-56.5		0.017	2.18
14	VIIIb	left	-23.0	-38.5	-53.5	12	0.01	2.42
15	I-IV	left	-0.5	-52.0	2.0	1	0.012	2.33
16	VIIIa	left	-20.0	-64.0	-59.5	35	0.013	2.29
17	VTA	left	-8.0	-17.5	-10.0	7	0.014	2.28
18	Crus I	right	23.5	-83.5	-22.0	10	0.015	2.24
19	Crus II	left	-9.5	-74.5	-34.0	15	0.016	2.23
20	VIIb	left	-18.5	-79.0	-53.5	11	0.016	2.22
21	VIIIa	right	29.5	-62.5	-53.5	8	0.017	2.19
22	VI	right	28.0	-59.5	-19.0	20	0.02	2.12
23	Extended cluster	right IX (18), vermal VIIIb (4), vermal VIIIa (2), right VIIIa (1)						
	IX	right	5.5	-58.0	-43.0	25	0.02	2.12
	VIIIa	right	7.0	-67.0	-46.0		0.042	1.76
24	V	left	-21.5	-49.0	-16.0	3	0.021	2.10
25	Crus II	right	4.0	-77.5	-37.0	8	0.022	2.07
26	white matter		5.5	-71.5	-11.5	1	0.024	2.03
27	VIIb	right	20.5	-77.5	-55.0	5	0.025	2.01
28	VI	right	20.5	-74.5	-17.5	1	0.027	1.97
29	VI	vermal	2.5	-62.5	-25.0	16	0.029	1.94
30	V	left	-21.5	-41.5	-23.5	6	0.03	1.93
31	VI	right	23.5	-71.5	-17.5	3	0.032	1.90
32	VIIIb	right	14.5	-56.5	-52.0	4	0.034	1.86
33	VIIb	left	-5.0	-71.5	-43.0	3	0.036	1.83
34	V	left	-20.0	-53.5	-16.0	6	0.037	1.82
35	VI	right	20.5	-59.5	-28.0	5	0.038	1.81
36	VIIb	right	13.0	-71.5	-53.5	9	0.039	1.80
37	V	left	-23.0	-34.0	-22.0	2	0.039	1.80
38	I-IV	right	10.0	-50.5	-5.5	1	0.041	1.78
39	VIIb	left	-42.5	-52.0	-53.5	1	0.041	1.77
40	DN	right	14.5	-62.5	-37.0	1	0.044	1.73
41	white matter		-11.0	-38.5	-35.5	1	0.046	1.72
42	VIIIb	left	-21.5	-52.0	-58.0	1	0.047	1.70
43	I-IV	right	5.5	-53.5	0.5	1	0.048	1.69

Table S9. fMRI activation clusters (p < 0.05, uncorrected) related to prediction, presentation and omission of the unconditioned stimulus (US) during acquisition and extinction training (Figure 4 and 5). Clusters were identified in the cerebellar cortex, deep cerebellar nuclei (DCN), and ventral tegmental area (VTA). Up to three local maxima per cluster are reported, separated by at least 8 mm. Coordinates are given in MNI space (x, y, z). Cluster size is reported as number of voxels (voxel volume = 3.375 mm^3). US: unconditioned stimulus; CS: conditioned stimulus; VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; MNI: Montreal Neurological Institute standard brain; t: t-statistic; punc: uncorrected p-value.

fMRI activations related to the unexpected omission of the US. TFCE and FWE corrected

Index	Location (lobule, DCN, VTA)	Side	MNI coordinates/mm			Cluster size (number of voxels)	p_{FWE}	TFCE t
			x	y	z			
Figure 6A: First 3 no US post CS+ > no US post CS- during extinction, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left Crus I (2540), left Crus II (2266), white matter (1957), left VI (1531), right VI (1403), right Crus I (1144), left VIIb (811), left DN (273), right V (240), right I-IV (208), left I-IV (185), right DN (154), left VIIIa (138), vermal VI (126), vermal VIIIa (115), left V (114), left IX (93), right Crus II (79), left VIIIb (57), vermal IX (29), right IX (29), left IN (20), left X (15), vermal VIIb (9), vermal Crus I (5), vermal Crus II (4), vermal VIIIb (4), right IN (4), vermal X (3), right VIIb (1), left FN (1)						
	Crus I	left	-38.0	-53.5	-35.5	13558	<0.001	3262
	VI	left	-20.0	-68.5	-26.5		0.001	3214
	VI	left	-24.5	-50.5	-31.0		0.001	3073
2	Extended cluster	right VIIb (193), right Crus II (113), right VIIIa (11)						
	Crus II	right	7.0	-79.0	-43.0	317	0.007	2053
	VIIb	right	22.0	-71.5	-52.0		0.013	1773
	VIIb	right	14.5	-76.0	-50.5		0.014	1746
3	I-IV	right	4.0	-52.0	0.5	2	0.018	1642
4	Extended cluster	left VTA (35), right VTA (22)						
	VTA	right	8.5	-17.5	-10.0	57	0.022	1562
	VTA	left	-8.0	-17.5	-10.0		0.022	1550
	VTA	left	-0.5	-16.0	-14.5		0.031	1413
5	Extended cluster	right VIIb (197), right Crus II (171), right VIIIa (146), white matter (31), right Crus I (24), right X (11), right VI (9)						
	VIIIa	right	35.5	-52.0	-52.0	589	0.022	1551
	Crus II	right	43.0	-49.0	-46.0		0.027	1470
	VIIb	right	35.5	-44.5	-43.0		0.028	1451
6	VI	right	20.5	-74.5	-17.5	1	0.023	1545
7	I-IV	right	5.5	-49.0	-1.0	3	0.023	1545
8	I-IV	left	-5.0	-49.0	-1.0	2	0.023	1545
9	VTA	right	1.0	-20.5	-8.5	5	0.031	1412
10	VTA	left	-0.5	-20.5	-17.5	2	0.036	1348
11	VI	left	-8.0	-67.0	-10.0	1	0.039	1315
12	Crus II	right	28.0	-83.5	-46.0	52	0.042	1281

Figure 6B: First 3 no US post CS+ > no US post CS- during recall, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left Crus I (1340), left VI (1298), left Crus II (599), left VIIb (142), vermal VI (132), white matter (55), left DN (20), left VIIIa (17), right VI (9), left V (6), vermal Crus I (4)						
	VI	left	-21.5	-71.5	-28.0	3622	0.001	2073
	VI	left	-29.0	-67.0	-26.5		0.002	2046
	VI	left	-8.0	-76.0	-26.5		0.002	1988
2	Extended cluster	vermal VIIIa (140), vermal VIIIb (138), vermal IX (137), left IX (55), left VIIIb (43), white matter (31), left VIIIa (9), vermal X (8), left DN (1)						
	IX	vermal	-0.5	-59.5	-35.5	562	0.021	1318
	VIIIb	vermal	-0.5	-65.5	-44.5		0.028	1245
	VIIIb	left	-9.5	-61.0	-43.0		0.028	1243
3	Extended cluster	right VI (344), white matter (159), right DN (46), right Crus I (30), right IN (8)						
	VI	right	34.0	-55.0	-29.5	587	0.021	1317
	VI	right	32.5	-55.0	-20.5		0.024	1291
	VI	right	20.5	-58.0	-28.0		0.024	1289
4	I-IV	right	14.5	-43.0	-11.5	43	0.024	1291
5	V	right	20.5	-53.5	-13.0	7	0.024	1288
6	Extended cluster	right VIIIa (295), right VIIb (230), right Crus II (94), right VIIIb (66), white matter (16), vermal VIIIa (2)						
	VIIb	right	19.0	-73.0	-49.0	703	0.027	1256
	VIIIa	right	23.5	-64.0	-49.0		0.027	1249
	VIIIa	right	32.5	-52.0	-52.0		0.03	1224
7	VTA	right	2.5	-17.5	-14.5	39	0.033	1192
8	Crus I	left	-36.5	-80.5	-25.0	1	0.035	1178
9	Extended cluster	left VIIb (194), left VIIIa (43), left Crus II (27), left VIIIb (1)						
	VIIb	left	-36.5	-47.5	-50.5	265	0.037	1164
	VIIb	left	-38.0	-55.0	-56.5		0.037	1162
	Crus II	left	-42.5	-62.5	-50.5		0.042	1130
10	VI	right	23.5	-77.5	-19.0	2	0.039	1151
11	V	right	8.5	-56.5	-4.0	2	0.039	1151
12	Extended cluster	right IX (86), white matter (32), right VIIIb (23), right X (17)						
	IX	right	13.0	-53.5	-46.0	158	0.039	1151
	VIIIb	right	19.0	-43.0	-49.0		0.045	1112
13	VIIIb	left	-17.0	-52.0	-50.5	41	0.04	1140
14	VI	right	13.0	-76.0	-16.0	3	0.042	1127
15	I-IV	right	4.0	-46.0	-5.5	52	0.044	1118
16	VTA	left	-8.0	-19.0	-11.5	1	0.044	1115
17	Extended cluster	left X (49), white matter (16), left VIIIb (12)						
	VIIIb	left	-24.5	-38.5	-47.5	77	0.045	1111
	X	left	-15.5	-40.0	-44.5		0.045	1110
18	VI	right	29.5	-47.5	-20.5	1	0.046	1102
19	VI	right	20.5	-71.5	-16.0	1	0.046	1102

20	V	right	23.5	-49.0	-16.0	1	0.046	1102
21	V	left	-17.0	-46.0	-13.0	1	0.048	1092
22	X	right	22.0	-38.5	-46.0	4	0.048	1090
Figure 6C: First 3 no US post CS+ > no US post CS- during reacquisition, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	right VTA (74), left VTA (71)						
	VTA	left	-2.0	-17.5	-7.0	145	0.001	7243
	VTA	right	8.5	-16.0	-11.5		0.001	6425
	VTA	left	-0.5	-20.5	-17.5		0.002	4461
2	Extended cluster	left Crus I (3033), left Crus II (2504), left VI (2143), right VI (2121), right Crus I (1081), left VIIb (982), white matter (896), right VIIb (878), left IX (695), right VIIIa (604), right Crus II (582), right IX (351), vermal IX (252), left VIIIb (246), right X (244), vermal VI (176), left X (175), right I-IV (153), right DN (152), left DN (151), left I-IV (148), right V (130), right VIIIb (121), left VIIIa (75), left V (61), vermal VIIIb (54), vermal VIIIa (47), vermal X (45), right IN (6), vermal Crus I (5), vermal Crus II (2), left FN (2), left IN (1)						
	Crus I	left	-9.5	-82.0	-28.0	18116	0.001	6274
	VI	left	-26.0	-65.5	-28.0		0.001	6189
	VI	left	-29.0	-59.5	-32.5		0.001	6171
3	white matter		-9.5	-38.5	-31.0	20	0.005	2394
4	white matter		-12.5	-37.0	-41.5	4	0.01	988
Figure 6D: First 3 no US post CS+ > no US post CS- during reextinction, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left Crus I (1030), left VI (807), left Crus II (63), white matter (1), vermal VI (1)						
	VI	left	-8.0	-76.0	-26.5	1902	0.004	2229
	Crus I	left	-15.5	-79.0	-25.0		0.005	2151
	VI	left	-30.5	-58.0	-31.0		0.007	2020
2	Extended cluster	left VTA (47), right VTA (34)						
	VTA	right	4.0	-14.5	-13.0	81	0.027	1432
	VTA	left	-5.0	-16.0	-13.0		0.028	1425
3	VTA	right	4.0	-14.5	-7.0	4	0.034	1340
4	Crus I	right	38.5	-53.5	-31.0	4	0.049	1194
Figure 7A: No US post CS+ x prediction error during extinction, t-test, TFCE, p < 0.05, FWE corr.								
1	Extended cluster	left Crus I (2341), left Crus II (1671), left VI (1220), white matter (529), left VIIb (337), left IX (269), vermal VIIIa (135), vermal IX (135), left DN (133), right I-IV (123), left I-IV (122), left VIIIb (64), left X (56), left VIIIa (44), vermal X (33), left V (30), vermal VIIIb (26), right V (23), vermal VI (15), left IN (10), vermal VIIb (9), vermal Crus I (6), vermal Crus II (4), right FN (4), left FN (2)						
	Crus I	left	-11.0	-79.0	-26.5	7341	0.002	2645
	VI	left	-23.0	-68.5	-28.0		0.003	2422
	Crus I	left	-20.0	-76.0	-31.0		0.003	2410
2	Extended cluster	left VTA (54), right VTA (52)						
	VTA	left	-5.0	-17.5	-13.0	106	0.011	1808
	VTA	right	5.5	-16.0	-14.5		0.013	1770

3	I-IV	right	5.5	-49.0	-1.0	3	0.021	1569
4	Extended cluster	right V (9), left V (2)						
	V	right	1.0	-62.5	-1.0	11	0.025	1496
	V	right	5.5	-67.0	-8.5		0.025	1484
5	I-IV	left	-5.0	-49.0	-1.0	3	0.026	1482
6	Extended cluster	right Crus I (600), right VI (246), right Crus II (16), vermal Crus II (1)						
	VI	right	37.0	-50.5	-31.0	863	0.034	1368
	Crus I	right	38.5	-76.0	-25.0		0.035	1356
	Crus I	right	13.0	-80.5	-26.5		0.036	1344
7	I-IV	right	4.0	-52.0	0.5	1	0.039	1315
8	Crus I	left	-45.5	-43.0	-31.0	1	0.043	1275
9	DN	right	16.0	-58.0	-34.0	136	0.044	1263
10	Crus I	right	23.5	-83.5	-22.0	2	0.046	1251
11	VI	right	16.0	-80.5	-19.0	4	0.046	1251
12	V	right	29.5	-44.5	-22.0	41	0.046	1248
	VI	right	35.5	-52.0	-22.0		0.049	1226
Figure 7B: No US post CS+ x prediction error during recall test, t-test, TFCE, p <0.05, FWE corr.								
1	Extended cluster	left VI (357), left Crus I (325), white matter (2)						
	VI	left	-33.5	-55.0	-32.5	684	0.003	1981
	VI	left	-29.0	-64.0	-29.5		0.006	1772
	VI	left	-38.0	-40.0	-35.5		0.043	1198
2	Extended cluster	left VI (56), vermal VI (11), left Crus I (9)						
	VI	vermal	-3.5	-79.0	-19.0	76	0.038	1231
	VI	left	-12.5	-80.5	-20.5		0.041	1213
	VI	left	-20.0	-77.5	-23.5		0.045	1180
3	Crus I	left	-14.0	-79.0	-29.5	60	0.043	1193
4	Crus II	left	-14.0	-79.0	-35.5	2	0.05	1151
Figure 7C: No US post CS+ x prediction error during reacquisition, t-test, TFCE, p < 0.05, FWE corr.								
	No significant voxels							
Figure 7D: No US post CS+ x prediction error during reextinction, t-test, TFCE, p < 0.05, FWE corr.								
	No significant voxels							

Table S10. fMRI activation clusters related to the unexpected omission of the unconditioned stimulus (US) during extinction training (Figure 6 and 7). Clusters were identified in the cerebellar cortex, deep cerebellar nuclei (DCN), and ventral tegmental area (VTA) using threshold-free cluster enhancement (TFCE) with family-wise error (FWE) correction ($p < 0.05$). Up to three local maxima per cluster are reported, separated by at least 8 mm. Coordinates are given in MNI space (x, y, z). Cluster size is reported as number of voxels (voxel volume = 3.375 mm^3). US: unconditioned stimulus; CS: conditioned stimulus; VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; MNI: Montreal Neurological Institute standard brain; TFCE t: threshold-free cluster-enhanced t-statistic; pFWE: family-wise error-corrected p-value.

fMRI activations related to the unexpected omission of the US. Uncorrected

Index	Location (lobule, DCN, VTA)	Side	MNI coordinates/mm			Cluster size (number of voxels)	p_{unc}	t
			x	y	z			
Figure 6A: First 3 no US post CS+ > no US post CS- during extinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (3015), white matter (2887), left Crus II (2640), right VI (1855), left VI (1794), right Crus I (1739), left VIIb (1040), right Crus II (838), right VIIb (580), right V (447), right I-IV (333), right VIIIa (327), left DN (289), left V (283), left I-IV (272), left VIIIa (258), left IX (255), vermal VI (232), vermal VIIIa (227), right DN (225), right VIIIb (166), left VIIIb (134), right IX (103), vermal IX (75), right X (57), vermal VIIIb (50), left X (41), vermal Crus II (29), left IN (22), vermal VIIb (17), right IN (11), vermal X (9), vermal Crus I (7), left FN (1)						
	VI	left	-20.0	-68.5	-26.5	20258	<0.001	6.92
	Crus I	left	-38.0	-53.5	-35.5		<0.001	6.77
	Crus II	left	-29.0	-77.5	-41.5		<0.001	6.08
2	Extended cluster	left VTA (53), right VTA (49)						
	VTA	left	-8.0	-17.5	-10.0	102	<0.001	3.82
	VTA	right	8.5	-17.5	-10.0		<0.001	3.74
	VTA	left	-0.5	-16.0	-14.5		0.004	2.77
3	I-IV	left	-5.0	-49.0	-1.0	6	0.012	2.34
4	I-IV	left	-0.5	-52.0	2.0	1	0.021	2.08
5	IX	right	5.5	-49.0	-59.5	1	0.027	1.97
6	VIIIb	right	20.5	-46.0	-56.5	6	0.029	1.94
7	Crus I	left	-33.5	-86.5	-32.5	2	0.031	1.91
8	white matter		22.0	-32.5	-37.0	5	0.031	1.91
9	V	left	-18.5	-47.5	-14.5	1	0.036	1.84
10	Crus II	left	-26.0	-89.5	-37.0	3	0.038	1.81
11	white matter		32.5	-62.5	-38.5	3	0.04	1.79
Figure 6B: First 3 no US post CS+ > no US post CS- during recall, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (1983), left VI (1831), left Crus II (1548), right VI (1164), white matter (1145), left VIIb (683), right VIIIa (557), right VIIb (554), right Crus II (384), left VIIIa (353), right I-IV (350), left VIIIb (343), right VIIIb (335), vermal VIIIa (294), vermal VI (273), right Crus I (205), left IX (197), vermal IX (197), right IX (189), vermal VIIIb (182), left X (162), right V (153), right DN (153), left I-IV (145), left DN (137), right X (121), left V (93), vermal X (25), right IN (21), vermal Crus II (19), vermal VIIb (9), vermal Crus I (7), left FN (3), left IN (1), right FN (1)						
	VI	left	-21.5	-71.5	-28.0	13817	<0.001	5.60
	VI	left	-8.0	-76.0	-26.5		<0.001	5.45
	VI	left	-35.0	-53.5	-31.0		<0.001	5.37

2	Extended cluster	right VTA (51), left VTA (47)						
	VTA	right	2.5	-17.5	-14.5	98	<0.001	4.75
	VTA	left	-8.0	-19.0	-11.5		<0.001	3.57
3	I-IV	right	13.0	-34.0	-22.0	26	0.002	3.09
4	V	left	-17.0	-46.0	-13.0	4	0.004	2.75
5	V	left	-14.0	-50.5	-10.0	6	0.01	2.44
6	Extended cluster	left I-IV (45)						
	I-IV	left	-8.0	-38.5	-20.5	45	0.012	2.34
	I-IV	left	-6.5	-41.5	-11.5		0.014	2.28
7	V	left	-3.5	-58.0	-13.0	13	0.013	2.31
8	Crus I	right	47.5	-53.5	-44.5	5	0.017	2.18
9	IX	right	5.5	-58.0	-55.0	8	0.022	2.07
10	VIIIa	right	10.0	-68.5	-53.5	6	0.032	1.90
11	VI	right	23.5	-71.5	-17.5	1	0.036	1.84
12	IX	left	-6.5	-56.5	-53.5	5	0.037	1.83
13	VI	right	10.0	-74.5	-14.5	1	0.038	1.81
14	V	left	-8.0	-58.0	-4.0	1	0.038	1.81
15	VI	left	-23.0	-62.5	-16.0	1	0.042	1.75
16	V	left	-0.5	-68.5	-8.5	1	0.046	1.71
17	white matter		-9.5	-41.5	-50.5	1	0.048	1.69
Figure 6C: First 3 no US post CS+ > no US post CS- reacquisition, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (2890), left Crus II (2337), left VI (2089), right VI (1996), right Crus I (986), left VIIb (932), right VIIb (789), white matter (747), left IX (650), right VIIIa (546), right Crus II (468), right IX (263), vermal IX (240), right X (240), left VIIIb (225), vermal VI (159), left X (157), left DN (132), right DN (121), left I-IV (106), right I-IV (90), right VIIIb (87), right V (70), left VIIIa (63), vermal VIIIa (41), vermal VIIIb (41), vermal X (37), left V (14), vermal Crus I (5), right IN (4), vermal Crus II (2), left FN (2)						
	Crus I	left	-38.0	-58.0	-31.0	16529	<0.001	9.00
	Crus I	left	-9.5	-82.0	-28.0		<0.001	8.92
	VI	left	-26.0	-65.5	-28.0		<0.001	8.24
2	Extended cluster	right VTA (72), left VTA (67)						
	VTA	left	-2.0	-17.5	-7.0	139	<0.001	6.35
	VTA	left	-0.5	-20.5	-17.5		<0.001	5.10
	VTA	right	8.5	-16.0	-11.5		<0.001	4.52
3	V	left	-0.5	-59.5	-1.0	126	<0.001	4.07
4	white matter		-9.5	-38.5	-31.0	17	0.002	3.08
5	I-IV	right	28.0	-31.0	-31.0	11	0.007	2.57
6	VI	right	35.5	-34.0	-35.5	3	0.029	1.94
7	I-IV	right	1.0	-55.0	-10.0	6	0.036	1.84
8	white matter		-12.5	-37.0	-41.5	1	0.036	1.84
9	V	left	-2.0	-61.0	-14.5	3	0.041	1.77
10	Crus II	right	14.5	-74.5	-35.5	3	0.042	1.76

11	VI	right	32.5	-34.0	-37.0	1	0.046	1.71
Figure 6D: First 3 no US post CS+ > no US post CS- reextinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (2250), left Crus II (1542), left VI (1426), left VIIb (606), white matter (460), left I-IV (322), right I-IV (241), left IX (220), vermal IX (147), left DN (118), right V (79), left VIIIa (73), left VIIIb (67), left V (56), right DN (44), vermal VI (29), right IN (18), vermal VIIIa (15), left IN (14), vermal Crus I (7), vermal VIIIb (6), vermal X (6), right IX (4), left FN (4), right FN (2), right VI (1)						
	VI	left	-8.0	-76.0	-26.5	7757	<0.001	6.40
	VI	left	-30.5	-58.0	-31.0		<0.001	5.84
	Crus I	left	-15.5	-79.0	-25.0		<0.001	5.76
2	Extended cluster	left VTA (73), right VTA (69)						
	VTA	left	-5.0	-16.0	-13.0	142	<0.001	5.04
	VTA	right	4.0	-14.5	-13.0		<0.001	5.03
3	Extended cluster	right VI (905), right Crus I (585), right VIIb (498), right VIIIa (433), right X (161), white matter (135), right VIIIb (40), vermal VI (34), right Crus II (29), right V (11), right IX (1)						
	Crus I	right	38.5	-53.5	-31.0	2832	<0.001	4.58
	Crus I	right	43.0	-61.0	-29.5		<0.001	3.54
	VIIb	right	14.5	-74.5	-47.5		0.001	3.51
4	white matter		-23.0	-47.5	-43.0	48	0.008	2.51
5	VIIIa	vermal	7.0	-64.0	-38.5	21	0.017	2.20
6	IX	right	11.5	-59.5	-47.5	27	0.017	2.19
7	white matter		-20.0	-40.0	-37.0	14	0.018	2.15
8	VI	vermal	4.0	-70.0	-16.0	13	0.025	2.01
9	VIIIb	right	11.5	-49.0	-61.0	5	0.028	1.96
10	white matter		-15.5	-34.0	-40.0	4	0.032	1.90
11	I-IV	left	-11.0	-41.5	-16.0	5	0.032	1.89
12	I-IV	left	-8.0	-50.5	-10.0	3	0.036	1.83
13	VI	left	-35.0	-35.5	-35.5	1	0.043	1.75
14	VI	vermal	-0.5	-68.5	-26.5	1	0.045	1.72
15	VI	vermal	1.0	-70.0	-28.0	1	0.046	1.71
16	Crus I	right	23.5	-73.0	-29.5	1	0.049	1.68
Figure 7A: No US post CS+ x prediction error during extinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (3530), left Crus II (2289), white matter (1583), left VI (1506), right Crus I (1298), right VI (798), left VIIb (606), left IX (536), left I-IV (254), vermal VIIIa (246), right I-IV (232), left DN (215), right DN (198), vermal IX (177), left VIIIb (173), right V (136), left X (127), vermal VI (126), right Crus II (85), left VIIIa (85), vermal VIIIb (69), left V (66), vermal X (60), vermal Crus II (55), vermal VIIb (41), right IX (20), left IN (18), right IN (16), right FN (8), left FN (7), vermal Crus I (6)						
	Crus I	left	-11.0	-79.0	-26.5	14566	<0.001	6.39
	VI	left	-23.0	-68.5	-28.0		<0.001	5.55
	VI	left	-33.5	-50.5	-32.5		<0.001	5.20

2	Extended cluster	right VTA (81), left VTA (71)						
	VTA	left	-5.0	-17.5	-13.0	152	<0.001	4.45
	VTA	right	5.5	-17.5	-14.5		<0.001	3.77
3	Extended cluster	right Crus II (410), right VIIb (296), right VIIIa (4)						
	Crus II	right	29.5	-82.0	-49.0	710	<0.001	4.04
	Crus II	right	8.5	-79.0	-43.0		<0.001	3.74
	VIIb	right	20.5	-71.5	-52.0		0.001	3.42
4	Extended cluster	right X (112), white matter (45), right VI (25), right VIIIa (19), right VIIIb (17), right VIIb (2)						
	VI	right	31.0	-41.5	-41.5	220	0.001	3.40
	X	right	20.5	-37.0	-49.0		0.006	2.61
	X	right	22.0	-32.5	-40.0		0.027	1.98
5	Extended cluster	right VIIb (107), right VIIIa (92), right Crus II (39), right VIIIb (8), white matter (4)						
	VIIb	right	41.5	-56.5	-53.5	250	0.001	3.16
	VIIIa	right	32.5	-52.0	-49.0		0.008	2.51
	VIIIb	right	28.0	-44.5	-55.0		0.022	2.07
6	Extended cluster	right V (31), left V (8)						
	V	right	1.0	-62.5	-1.0	39	0.005	2.68
	V	right	5.5	-67.0	-8.5		0.006	2.60
7	Extended cluster	right I-IV (23)						
	I-IV	right	5.5	-49.0	-1.0	23	0.006	2.60
	I-IV	right	2.5	-56.5	0.5		0.019	2.14
8	I-IV	left	-5.0	-49.0	-1.0	7	0.008	2.49
9	Crus II	right	31.0	-68.5	-41.5	83	0.01	2.44
10	Crus I	left	-50.0	-70.0	-40.0	12	0.014	2.28
11	IX	right	8.5	-58.0	-53.5	18	0.014	2.27
12	V	left	-26.0	-31.0	-29.5	6	0.019	2.14
13	white matter		5.5	-71.5	-11.5	9	0.024	2.03
14	VI	right	38.5	-38.5	-29.5	5	0.028	1.95
15	VI	right	26.5	-58.0	-29.5	9	0.033	1.87
16	Crus II	right	46.0	-47.5	-46.0	2	0.034	1.86
17	VIIIa	left	-32.0	-47.5	-56.5	1	0.038	1.82
18	I-IV	left	-8.0	-46.0	-4.0	2	0.042	1.76
19	VI	left	-38.0	-37.0	-31.0	1	0.042	1.76
20	VI	right	41.5	-41.5	-32.5	3	0.042	1.75
Figure 7B: No US post CS+ x prediction error during recall, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (1637), left VI (1505), right VI (992), white matter (767), left Crus II (418), vermal VI (364), left VIIIb (322), vermal VIIIa (308), left X (221), left VIIIa (190), vermal IX (186), left IX (175), vermal VIIIb (154), right DN (151), right X (140), right Crus I (112), left I-IV (110), left DN (92), right I-IV (67), right Crus II (64), right V (62), vermal Crus II (62), vermal VIIb (55), left V (53), right VIIIb (49), right IX (41), vermal						

		X (38), right VIIb (25), right VIIIa (18), left VIIb (10), left IN (10), vermal Crus I (4), right IN (1), right FN (1)						
	VI	left	-33.5	-55.0	-32.5	8404	<0.001	7.24
	VI	left	-29.0	-64.0	-29.5		<0.001	5.98
	VIIIb	vermal	-3.5	-64.0	-41.5		<0.001	4.74
2	Extended cluster	left VTA (51), right VTA (32)						
	VTA	left	-0.5	-16.0	-13.0	83	<0.001	4.50
	VTA	right	7.0	-20.5	-14.5		0.019	2.13
3	Extended cluster	left Crus II (535), left VIIb (170), left VIIIa (16), left Crus I (9), white matter (1)						
	Crus II	left	-27.5	-82.0	-43.0	731	<0.001	3.84
	Crus II	left	-23.0	-76.0	-46.0		0.006	2.60
	VIIIa	left	-18.5	-70.0	-55.0		0.015	2.23
4	Extended cluster	right VIIb (278), right VIIIa (273), right VIIIb (135), right Crus II (125), white matter (93), right Crus I (2)						
	VIIIb	right	29.5	-46.0	-53.5	906	<0.001	3.83
	VIIb	right	22.0	-67.0	-44.5		0.001	3.43
	VIIb	right	40.0	-44.5	-50.5		0.002	3.04
5	IX	left	-6.5	-58.0	-53.5	39	0.002	2.97
6	Extended cluster	right I-IV (29), left V (6), right V (6), left I-IV (3)						
	I-IV	right	4.0	-52.0	0.5	44	0.003	2.84
	V	left	-0.5	-61.0	-1.0		0.01	2.43
7	Crus II	left	-33.5	-59.5	-44.5	62	0.004	2.82
8	Extended cluster	left Crus II (89), left VIIb (56)						
	Crus II	left	-45.5	-52.0	-50.5	145	0.004	2.77
	Crus II	left	-39.5	-74.5	-50.5		0.033	1.88
	Crus II	left	-42.5	-67.0	-50.5		0.038	1.81
9	VIIb	left	-3.5	-76.0	-49.0	2	0.009	2.47
10	Crus I	right	34.0	-79.0	-22.0	21	0.012	2.33
11	I-IV	right	13.0	-35.5	-23.5	11	0.012	2.33
12	Crus I	right	43.0	-55.0	-34.0	34	0.014	2.28
13	I-IV	right	2.5	-58.0	-5.5	29	0.017	2.18
14	VIIIb	right	13.0	-47.5	-64.0	5	0.018	2.17
15	Crus I	right	43.0	-64.0	-34.0	21	0.018	2.15
16	Crus I	right	49.0	-73.0	-38.5	1	0.033	1.89
17	VI	right	37.0	-43.0	-25.0	1	0.036	1.83
18	VI	right	13.0	-73.0	-25.0	2	0.037	1.82
19	VIIIa	left	-9.5	-65.5	-56.5	1	0.037	1.82
20	Crus II	right	41.5	-68.5	-52.0	5	0.038	1.81
21	I-IV	left	-3.5	-50.5	-8.5	4	0.039	1.80
22	VIIb	right	34.0	-65.5	-56.5	1	0.041	1.77
23	Crus I	left	-21.5	-86.5	-25.0	2	0.041	1.77
24	IX	right	5.5	-58.0	-55.0	1	0.043	1.75
25	V	left	-6.5	-58.0	-13.0	1	0.047	1.70
26	VIIIa	right	17.5	-64.0	-58.0	1	0.048	1.69

Figure 7C: No US post CS+ x prediction error during reacquisition, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (1123), left VI (925), left Crus II (863), left VIIb (381), white matter (53), left DN (5), left VIIIa (4)						
	Crus I	left	-27.5	-68.5	-31.0	3354	<0.001	3.91
	VI	left	-30.5	-56.5	-32.5		<0.001	3.77
	Crus I	left	-35.0	-49.0	-35.5		0.001	3.52
2	Extended cluster	right VIIb (287), right Crus II (154), right VIIIa (154)						
	VIIIa	right	32.5	-61.0	-53.5	595	<0.001	3.70
	VIIb	right	17.5	-74.5	-46.0		<0.001	3.59
	VIIb	right	31.0	-71.5	-55.0		0.002	3.01
3	Extended cluster	right VTA (32), left VTA (9)						
	VTA	right	5.5	-13.0	-11.5	41	<0.001	3.61
	VTA	left	-0.5	-20.5	-17.5		<0.001	3.60
4	Extended cluster	right VI (270), right Crus I (124)						
	VI	right	37.0	-52.0	-31.0	394	0.001	3.35
	VI	right	32.5	-64.0	-25.0		0.039	1.80
5	Extended cluster	left IX (175), vermal IX (77), right IX (70), left VIIIb (67), white matter (11), vermal VIIIa (4), vermal VIIIb (4), vermal X (2), left VIIIa (1)						
	IX	right	2.5	-58.0	-56.5	411	0.001	3.29
	white matter		-9.5	-58.0	-41.5		0.005	2.71
	VIIIb	left	-9.5	-62.5	-52.0		0.008	2.51
6	X	left	-24.5	-34.0	-46.0	32	0.001	3.23
7	white matter		-24.5	-41.5	-40.0	77	0.003	2.91
8	white matter		26.5	-43.0	-43.0	52	0.005	2.68
9	VTA	left	-5.0	-13.0	-11.5	11	0.012	2.33
10	DN	left	-15.5	-50.5	-35.5	13	0.018	2.15
11	IX	right	5.5	-49.0	-59.5	3	0.024	2.02
12	white matter		-9.5	-40.0	-38.5	4	0.027	1.98
13	X	right	20.5	-35.5	-46.0	9	0.028	1.96
14	VTA	left	-3.5	-16.0	-14.5	3	0.031	1.91
15	Crus I	right	47.5	-64.0	-26.5	5	0.033	1.88
16	VTA	left	-8.0	-16.0	-10.0	2	0.036	1.84
17	Crus II	right	44.5	-71.5	-44.5	3	0.038	1.81
18	white matter		-9.5	-38.5	-31.0	2	0.039	1.80
19	Crus II	right	11.5	-73.0	-35.5	1	0.039	1.80
20	DN	right	17.5	-59.5	-34.0	2	0.039	1.80
21	VIIIb	right	10.0	-44.5	-61.0	4	0.039	1.79
22	Crus I	left	-42.5	-41.5	-41.5	2	0.043	1.75
23	VIIIb	right	14.5	-55.0	-56.5	2	0.048	1.70
24	VTA	right	1.0	-20.5	-7.0	1	0.048	1.70

Figure 7D: No US post CS+ x prediction error during recall, t-test, p < 0.05, uncorrected								
1	Extended cluster	right VIIb (160), right Crus II (67), right VIIIa (31)						
	VIIb	right	16.0	-74.5	-47.5	258	0.002	3.12
	VIIb	right	31.0	-71.5	-53.5		0.014	2.28
	VIIIa	right	32.5	-61.0	-53.5		0.017	2.19
2	VIIIb	right	11.5	-46.0	-61.0	41	0.003	2.92
3	Extended cluster	left Crus II (216), left VIIb (179), white matter (46), left VIIIa (8), left Crus I (5)						
	VIIb	left	-32.0	-62.5	-49.0	454	0.005	2.70
	Crus II	left	-27.5	-71.5	-47.5		0.005	2.67
	VIIb	left	-15.5	-73.0	-49.0		0.007	2.54
4	VTA	left	-0.5	-20.5	-17.5	12	0.005	2.66
5	white matter		-24.5	-41.5	-40.0	63	0.006	2.61
6	Extended cluster	left IX (26), right IX (26), left VIIIb (6), right VIIIb (3)						
	IX	left	-5.0	-58.0	-59.5	61	0.008	2.51
	IX	right	2.5	-58.0	-56.5		0.014	2.26
	IX	right	7.0	-56.5	-64.0		0.026	1.99
7	X	left	-26.0	-35.5	-47.5	8	0.012	2.35
8	Crus I	left	-38.0	-62.5	-31.0	37	0.016	2.22
9	VI	left	-35.0	-35.5	-35.5	25	0.018	2.17
10	VIIIb	right	23.5	-49.0	-50.5	4	0.023	2.06
11	VIIIa	vermal	-0.5	-73.0	-40.0	10	0.023	2.06
12	I-IV	left	-5.0	-55.0	-16.0	4	0.023	2.06
13	Crus II	left	-21.5	-86.5	-46.0	21	0.023	2.04
14	I-IV	left	-9.5	-37.0	-22.0	11	0.024	2.03
15	VI	left	-6.5	-73.0	-26.5	20	0.026	2.00
16	Extended cluster	white matter (16), left DN (9), left VI (1)						
	white matter		-24.5	-55.0	-35.5	26	0.027	1.97
	white matter		-17.0	-58.0	-32.5		0.038	1.81
17	VI	left	-32.0	-55.0	-34.0	21	0.028	1.96
18	Crus II	right	44.5	-73.0	-43.0	7	0.028	1.96
19	white matter		-9.5	-40.0	-40.0	1	0.029	1.95
20	Crus I	left	-15.5	-80.5	-25.0	9	0.03	1.92
21	Crus I	left	-36.5	-49.0	-37.0	9	0.033	1.89
22	white matter		-12.5	-50.5	-25.0	7	0.034	1.86
23	VIIIb	left	-8.0	-61.0	-40.0	2	0.035	1.85
24	Crus I	right	40.0	-52.0	-31.0	10	0.036	1.83
25	VIIIa	vermal	5.5	-67.0	-38.5	5	0.037	1.83
26	V	right	4.0	-58.0	-26.5	3	0.038	1.80
27	VTA	left	-5.0	-13.0	-11.5	1	0.04	1.78

28	white matter		-11.0	-68.5	-32.5	9	0.04	1.78
29	VTA	right	4.0	-14.5	-13.0	1	0.042	1.76
30	white matter		-15.5	-40.0	-40.0	3	0.042	1.76
31	IX	vermal	-3.5	-58.0	-38.5	3	0.042	1.75
32	VIIb	right	40.0	-68.5	-55.0	1	0.046	1.72
33	Crus II	right	40.0	-67.0	-43.0	1	0.047	1.70
34	Crus II	right	35.5	-71.5	-47.5	1	0.049	1.68

Table S11. fMRI activation clusters ($p < 0.05$, uncorrected) related to the unexpected omission of the unconditioned stimulus (US) during extinction training (Figure 6 and 7). Clusters were identified in the cerebellar cortex, deep cerebellar nuclei (DCN), and ventral tegmental area (VTA). Up to three local maxima per cluster are reported, separated by at least 8 mm. Coordinates are given in MNI space (x, y, z). Cluster size is reported as number of voxels (voxel volume = 3.375 mm^3). US: unconditioned stimulus; CS: conditioned stimulus; VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; MNI: Montreal Neurological Institute standard brain; t: t-statistic; punc: uncorrected p-value.

fMRI activations related to PPI connectivity during unexpected US omissions with a VTA seed. Uncorrected

Index	Location (lobule, DCN)	Side	MNI coordinates/mm			Cluster size (number of voxels)	p_{unc}	t
			x	y	z			
Figure 8A: No US post CS+ x prediction error PPI positive connectivity with VTA during extinction, t-test, p < 0.05, uncorrected								
1	VIIIa	left	-24.5	-68.5	-59.5	107	<0.001	4.14
2	VI	left	-29.0	-59.5	-25.0	48	0.002	3.10
3	VI	left	-35.0	-35.5	-34.0	12	0.002	3.04
4	VIIIa	right	10.0	-70.0	-55.0	10	0.007	2.58
5	VIIb	left	-14.0	-73.0	-53.5	12	0.008	2.51
6	VIIIa	vermal	-2.0	-68.5	-44.5	6	0.009	2.47
7	VIIIb	right	17.5	-40.0	-53.5	9	0.009	2.45
8	V	right	19.0	-47.5	-20.5	8	0.01	2.42
9	VIIIb	right	17.5	-59.5	-59.5	8	0.011	2.39
10	V	right	10.0	-56.5	-19.0	17	0.012	2.36
11	VI	vermal	-3.5	-67.0	-23.5	6	0.013	2.32
12	VI	right	35.5	-34.0	-35.5	5	0.014	2.28
14	Crus II	left	-30.5	-79.0	-50.5	2	0.019	2.13
15	Crus I	left	-23.0	-89.5	-32.5	1	0.021	2.09
16	Crus II	right	16.0	-83.5	-35.5	1	0.021	2.09
17	Crus I	right	32.5	-86.5	-34.0	1	0.022	2.08
18	Crus II	left	-47.0	-53.5	-49.0	3	0.022	2.08
19	Crus II	vermal	4.0	-74.5	-32.5	1	0.024	2.03
20	Crus I	right	29.5	-88.0	-34.0	1	0.025	2.02

Figure 8A: No US post CS+ x prediction error PPI negative connectivity with VTA during extinction, t-test, p < 0.05, uncorrected								
1	white matter		29.5	-65.5	-44.5	84	0.001	3.36
2	Extended cluster	white matter (216), left VI (12), left VIIIa (1)						
	white matter		-20.0	-38.5	-37.0	229	0.002	3.14
	white matter		-23.0	-52.0	-40.0		0.003	2.91
	white matter		-26.0	-43.0	-40.0		0.008	2.51
3	Extended cluster	white matter (37), right IX (29), vermal X (15), right IN (10), vermal IX (8), vermal VIIIa (2)						
	IN	right	7.0	-58.0	-31.0	101	0.002	3.12
	X	vermal	5.5	-49.0	-34.0		0.007	2.57
4	DN	left	-12.5	-50.5	-32.5	82	0.002	3.07
5	Extended cluster	white matter (124), right X (4), right IX (1)						
	white matter		11.5	-41.5	-40.0	129	0.002	3.04
	white matter		23.5	-52.0	-43.0		0.011	2.39
6	Extended cluster	white matter (127), left Crus II (38), left Crus I (15), left DN (9)						
	white matter		-17.0	-70.0	-40.0	189	0.002	2.99
	white matter		-18.5	-61.0	-38.5		0.006	2.60
7	IX	right	5.5	-56.5	-64.0	10	0.004	2.83
8	V	right	8.5	-55.0	-4.0	38	0.005	2.70
9	Crus I	right	47.5	-67.0	-34.0	32	0.006	2.60
10	Crus II	right	37.0	-58.0	-43.0	19	0.006	2.60
11	Crus I	left	-36.5	-58.0	-31.0	34	0.006	2.60
12	I-IV	left	-20.0	-32.5	-22.0	2	0.011	2.39
13	Crus I	left	-26.0	-79.0	-26.5	10	0.011	2.37
14	Crus I	right	17.5	-85.0	-25.0	35	0.011	2.36
15	VIIIb	vermal	4.0	-61.0	-40.0	2	0.012	2.33
16	Crus I	left	-14.0	-86.5	-23.5	23	0.013	2.31
17	VIIIa	right	34.0	-49.0	-46.0	8	0.014	2.28
18	Crus II	left	-30.5	-67.0	-47.5	9	0.014	2.27
19	I-IV	right	19.0	-34.0	-19.0	1	0.015	2.25
20	white matter		23.5	-62.5	-40.0	13	0.015	2.24
21	Crus II	left	-38.0	-70.0	-43.0	3	0.016	2.22
22	VIIIb	right	10.0	-62.5	-55.0	4	0.016	2.20
23	Crus I	left	-42.5	-61.0	-37.0	4	0.017	2.20
24	IX	right	2.5	-55.0	-56.5	4	0.017	2.18
25	Crus II	left	-3.5	-88.0	-29.5	1	0.018	2.17

26	white matter		-3.5	-55.0	-29.5	1	0.018	2.16
27	VIIIb	left	-18.5	-53.5	-50.5	5	0.019	2.14
28	VI	right	28.0	-43.0	-32.5	3	0.019	2.13
29	VIIb	right	22.0	-77.5	-55.0	1	0.02	2.11
30	IX	right	4.0	-64.0	-53.5	2	0.02	2.11
31	Crus I	left	-45.5	-59.5	-25.0	6	0.02	2.11
32	Crus II	right	10.0	-80.5	-50.5	1	0.021	2.10
33	VIIb	left	-5.0	-74.5	-50.5	1	0.022	2.06
34	Crus I	left	-45.5	-65.5	-34.0	1	0.023	2.05
35	Crus II	right	13.0	-82.0	-50.5	1	0.024	2.04
36	white matter		29.5	-50.5	-41.5	1	0.024	2.04
37	IX	vermal	2.5	-50.5	-40.0	1	0.025	2.02
Figure 8A: No US post CS+ x prediction error PPI positive connectivity with VTA during recall, t-test, p <0.05, uncorrected								
1	Extended cluster	left VI (203), left Crus I (65), left V (52)						
	V	left	-21.5	-50.5	-20.5	320	<0.001	4.07
	Crus I	left	-39.5	-59.5	-26.5		<0.001	3.58
	VI	left	-32.0	-58.0	-23.5		0.001	3.21
2	Extended cluster	left VIIIa (94), left VIIIb (69), left IX (63), vermal VIIIa (25), left VIIb (15), vermal VIIIb (5), white matter (2), left X (2)						
	VIIIa	left	-5.0	-68.5	-46.0	275	0.001	3.53
	VIIIb	left	-14.0	-44.5	-49.0		0.004	2.81
	VIIIa	left	-11.0	-65.5	-50.5		0.006	2.65
3	Extended cluster	left I-IV (60), right I-IV (54)						
	I-IV	left	-0.5	-43.0	-14.5	114	0.001	3.41
	I-IV	right	4.0	-46.0	-23.5		0.001	3.26
	I-IV	left	-6.5	-43.0	-23.5		0.013	2.31
4	Extended cluster	right VI (134), right Crus I (67), white matter (11)						
	VI	right	25.0	-68.5	-29.5	212	0.001	3.29
	Crus I	right	34.0	-68.5	-28.0		0.003	2.90
	VI	right	19.0	-61.0	-29.5		0.004	2.75
5	I-IV	left	-15.5	-37.0	-25.0	69	0.001	3.28
6	Extended cluster	vermal VI (104), left VI (28), right VI (17), left Crus II (12), vermal Crus II (12), vermal Crus I (5), right Crus I (3), left Crus I (1)						
	Crus II	vermal	-0.5	-79.0	-28.0	182	0.001	3.24
	VI	right	7.0	-71.5	-20.5		0.004	2.78
	VI	left	-6.5	-73.0	-20.5		0.004	2.77
7	Extended cluster	right IX (107), right VIIIa (70), vermal VIIIb (41), vermal VIIIa (39), vermal IX (36), right VIIIb (30), right VIIb (7), white matter (4)						
	VIIIa	right	13.0	-67.0	-47.5	334	0.002	3.08
	VIIIb	vermal	7.0	-62.5	-37.0		0.003	2.90
	IX	right	11.5	-56.5	-49.0		0.003	2.89

8	Extended cluster	right VI (198), right V (155), right I-IV (26)						
	V	right	20.5	-46.0	-17.5	379	0.002	3.01
	VI	right	29.5	-55.0	-28.0		0.003	2.94
	V	right	7.0	-56.5	-7.0		0.003	2.92
9	VIIb	right	41.5	-53.5	-49.0	46	0.003	2.86
10	IX	left	-2.0	-55.0	-46.0	76	0.005	2.68
11	VI	left	-24.5	-73.0	-22.0	49	0.005	2.68
12	Extended cluster	left Crus I (59), white matter (5)						
	Crus I	left	-39.5	-76.0	-34.0	64	0.006	2.64
13	VI	left	-32.0	-44.5	-29.5	25	0.006	2.61
14	Crus II	right	22.0	-86.5	-41.5	17	0.006	2.59
15	VIIIa	right	31.0	-44.5	-49.0	14	0.007	2.59
16	VI	left	-17.0	-62.5	-29.5	7	0.007	2.55
17	VI	right	34.0	-41.5	-28.0	24	0.007	2.54
18	V	right	4.0	-64.0	-7.0	31	0.008	2.50
19	Crus II	left	-47.0	-55.0	-49.0	20	0.009	2.45
20	Crus I	left	-33.5	-74.5	-25.0	26	0.01	2.40
21	VIIIb	right	17.5	-56.5	-55.0	12	0.01	2.40
22	V	left	-9.5	-50.5	-8.5	10	0.012	2.35
23	I-IV	left	-8.0	-43.0	-7.0	6	0.012	2.33
24	VIIIb	left	-23.0	-47.5	-59.5	5	0.012	2.32
25	X	left	-21.5	-34.0	-46.0	26	0.013	2.31
26	Crus II	right	28.0	-80.5	-40.0	8	0.014	2.27
27	Crus I	right	41.5	-67.0	-37.0	5	0.014	2.26
28	Crus I	left	-20.0	-83.5	-29.5	8	0.016	2.22
29	Crus I	left	-12.5	-86.5	-28.0	5	0.016	2.22
30	Crus II	right	38.5	-82.0	-40.0	1	0.016	2.22
31	Crus II	left	-36.5	-40.0	-41.5	2	0.016	2.22
32	VI	left	-6.5	-67.0	-10.0	11	0.016	2.21
33	VI	right	38.5	-61.0	-22.0	1	0.016	2.20
34	VI	vermal	1.0	-64.0	-20.5	4	0.016	2.20
35	VIIIb	right	20.5	-43.0	-47.5	6	0.017	2.19
36	VI	left	-15.5	-67.0	-26.5	3	0.017	2.19
37	Crus I	left	-33.5	-70.0	-32.5	4	0.017	2.19
38	VIIb	left	-36.5	-41.5	-47.5	4	0.017	2.18
39	VIIIa	left	-30.5	-47.5	-52.0	3	0.018	2.15
40	X	left	-14.0	-38.5	-46.0	1	0.019	2.14
41	I-IV	right	7.0	-47.5	-7.0	3	0.02	2.12
42	V	left	-11.0	-59.5	-19.0	2	0.022	2.08
43	VI	left	-24.5	-64.0	-28.0	1	0.022	2.07
44	VI	right	35.5	-59.5	-22.0	1	0.023	2.05
45	VI	right	16.0	-74.5	-20.5	1	0.023	2.05
46	VI	left	-36.5	-41.5	-28.0	1	0.023	2.05
47	VIIIa	right	22.0	-61.0	-59.5	1	0.023	2.05
48	I-IV	right	13.0	-40.0	-28.0	1	0.024	2.03

49	white matter		-9.5	-59.5	-38.5	1	0.024	2.03
50	Crus II	left	-36.5	-80.5	-43.0	1	0.024	2.02
51	V	right	10.0	-61.0	-19.0	1	0.025	2.02
Figure 8A: No US post CS+ x prediction error PPI negative connectivity with VTA during recall, t-test, p < 0.05, uncorrected								
1	white matter		-20.0	-34.0	-32.5	17	0.002	2.98
2	white matter		-8.0	-68.5	-34.0	15	0.003	2.84
3	VIIb	right	40.0	-64.0	-53.5	24	0.004	2.83
4	Crus II	right	4.0	-80.5	-44.5	13	0.004	2.77
5	Crus II	left	-8.0	-91.0	-35.5	3	0.011	2.40
6	VIIb	left	-39.5	-59.5	-58.0	33	0.011	2.38
7	white matter		26.5	-44.5	-41.5	9	0.011	2.37
8	Crus II	right	8.5	-89.5	-34.0	2	0.012	2.34
9	VI	left	-27.5	-38.5	-32.5	5	0.015	2.25
10	Crus I	right	28.0	-83.5	-26.5	1	0.02	2.11
11	I-IV	right	20.5	-32.5	-22.0	1	0.021	2.09
12	Crus I	right	44.5	-77.5	-31.0	2	0.022	2.08
13	VIIb	right	26.5	-68.5	-55.0	1	0.023	2.05
14	white matter		-15.5	-34.0	-37.0	1	0.023	2.05
15	Crus I	right	44.5	-59.5	-29.5	1	0.024	2.03
Figure 8A: No US post CS+ x prediction error PPI positive connectivity with VTA during reacquisition, t-test, p < 0.05, uncorrected								
1	I-IV	left	-18.5	-32.5	-22.0	52	<0.001	3.95
2	VIIb	left	-5.0	-67.0	-32.5	82	0.001	3.26
3	Extended cluster	right Crus II (108), right VIIb (9)						
	Crus II	right	25.0	-80.5	-44.5	117	0.001	3.16
	Crus II	right	28.0	-71.5	-47.5		0.014	2.29
4	Extended cluster	left Crus I (96)						
	Crus I	left	-45.5	-58.0	-35.5	96	0.002	3.06
	Crus I	left	-39.5	-65.5	-26.5		0.015	2.25
5	VIIb	left	-39.5	-59.5	-55.0	25	0.003	2.88
6	white matter		7.0	-62.5	-28.0	34	0.003	2.88
7	Crus II	right	7.0	-82.0	-46.0	9	0.004	2.75
8	VI	left	-21.5	-62.5	-28.0	93	0.005	2.71
9	Extended cluster	right IX (20), left IX (12), left VIIIb (1)						
	IX	right	2.5	-62.5	-56.5	33	0.005	2.68
	IX	left	-6.5	-59.5	-58.0		0.014	2.29
10	V	right	16.0	-47.5	-23.5	86	0.005	2.67
11	I-IV	left	-11.0	-41.5	-10.0	15	0.006	2.65
12	VI	right	34.0	-46.0	-34.0	31	0.007	2.58

13	I-IV	right	17.5	-32.5	-20.5	25	0.007	2.57
14	white matter		17.5	-47.5	-41.5	11	0.007	2.56
15	VIIIa	right	26.5	-58.0	-61.0	6	0.007	2.54
16	VIIIa	right	11.5	-70.0	-56.5	3	0.009	2.48
17	white matter		31.0	-61.0	-43.0	9	0.009	2.47
18	white matter		25.0	-67.0	-41.5	37	0.009	2.46
19	Crus I	right	41.5	-70.0	-26.5	8	0.011	2.40
20	white matter		41.5	-55.0	-38.5	31	0.011	2.37
21	VI	vermal	2.5	-77.5	-16.0	3	0.012	2.34
22	VIIIb	right	25.0	-52.0	-58.0	5	0.013	2.32
23	I-IV	right	25.0	-32.5	-34.0	10	0.013	2.32
24	VI	right	19.0	-65.5	-28.0	4	0.014	2.28
25	I-IV	right	2.5	-43.0	-22.0	16	0.014	2.27
26	I-IV	right	11.5	-43.0	-16.0	14	0.015	2.25
27	VI	right	31.0	-61.0	-22.0	2	0.016	2.23
28	Crus II	left	-20.0	-82.0	-50.5	6	0.016	2.22
29	VI	right	28.0	-49.0	-20.5	4	0.016	2.22
30	Crus I	right	40.0	-53.5	-29.5	5	0.016	2.21
31	Crus II	left	-45.5	-65.5	-52.0	2	0.016	2.21
32	VI	right	35.5	-56.5	-23.5	6	0.016	2.21
33	Crus II	left	-5.0	-83.5	-43.0	4	0.018	2.17
34	Crus II	left	-33.5	-82.0	-47.5	1	0.018	2.16
35	I-IV	right	11.5	-46.0	-10.0	8	0.018	2.15
36	VI	left	-32.0	-67.0	-26.5	10	0.019	2.15
37	VI	left	-27.5	-53.5	-32.5	1	0.02	2.12
38	VIIb	right	20.5	-70.0	-52.0	1	0.022	2.07
39	VIIb	left	-38.0	-47.5	-50.5	1	0.023	2.05
40	VIIIb	right	28.0	-38.5	-50.5	1	0.023	2.05
41	VI	right	29.5	-74.5	-20.5	1	0.023	2.05
42	Crus II	left	-14.0	-80.5	-49.0	1	0.023	2.05
43	Crus I	left	-50.0	-61.0	-26.5	1	0.023	2.05
44	DN	left	-14.0	-55.0	-32.5	1	0.024	2.03
45	VI	left	-35.0	-53.5	-28.0	1	0.024	2.03
46	I-IV	right	14.5	-38.5	-25.0	1	0.025	2.02
Figure 8A: No US post CS+ x prediction error PPI negative connectivity with VTA during reacquisition, t-test, p <0.05, uncorrected								
1	white matter		-11.0	-47.5	-23.5	52	0.001	3.46
2	I-IV	right	4.0	-52.0	-13.0	27	0.001	3.15
3	Crus I	right	43.0	-47.5	-28.0	5	0.002	3.04
4	VIIb	right	44.5	-52.0	-49.0	15	0.002	2.98
5	VIIIb	left	-8.0	-44.5	-58.0	27	0.002	2.98
6	Extended cluster	white matter (66), left Crus I (1), left Crus II (1)						

	white matter		-30.5	-58.0	-40.0	68	0.002	2.97
	white matter		-21.5	-50.5	-37.0		0.004	2.80
7	white matter		8.5	-46.0	-31.0	12	0.003	2.91
8	VI	right	34.0	-40.0	-26.5	7	0.009	2.48
9	Crus II	right	11.5	-83.5	-40.0	12	0.009	2.47
10	Crus II	left	-12.5	-85.0	-40.0	9	0.009	2.46
11	white matter		-15.5	-40.0	-31.0	20	0.012	2.35
12	V	left	-21.5	-37.0	-29.5	9	0.014	2.27
13	VIIIb	right	16.0	-49.0	-52.0	12	0.014	2.26
14	VIIb	right	38.5	-44.5	-44.5	5	0.015	2.25
15	VIIIa	right	8.5	-70.0	-50.5	2	0.016	2.20
16	Crus I	left	-48.5	-47.5	-40.0	1	0.019	2.13
17	I-IV	left	-2.0	-46.0	-16.0	1	0.023	2.04
Figure 8A: No US post CS+ x prediction error PPI positive connectivity with VTA during reextinction, t-test, p < 0.05, uncorrected								
1	white matter		22.0	-32.5	-34.0	20	0.002	2.97
2	VIIIa	right	11.5	-70.0	-56.5	11	0.003	2.87
3	I-IV	left	-18.5	-32.5	-23.5	17	0.004	2.78
4	VIIIb	left	-18.5	-46.0	-61.0	21	0.007	2.57
5	Extended cluster	left VIIb (28), left VIIIa (11)						
	VIIb	left	-38.0	-53.5	-58.0	39	0.008	2.53
	VIIb	left	-38.0	-61.0	-53.5		0.017	2.20
	VIIIa	left	-29.0	-49.0	-58.0		0.018	2.16
6	VIIIb	right	11.5	-50.5	-61.0	9	0.008	2.52
7	I-IV	right	2.5	-43.0	-22.0	21	0.008	2.48
8	VI	left	-23.0	-59.5	-29.5	26	0.009	2.48
9	X	left	-15.5	-37.0	-49.0	3	0.011	2.37
10	VIIIa	left	-29.0	-40.0	-50.5	11	0.012	2.34
11	VIIb	right	41.5	-59.5	-55.0	17	0.012	2.33
12	Crus I	left	-50.0	-68.5	-41.5	2	0.014	2.28
13	Crus I	left	-47.0	-58.0	-34.0	8	0.014	2.27
14	VI	left	-36.5	-53.5	-28.0	9	0.014	2.26
15	VIIIb	right	14.5	-56.5	-61.0	7	0.016	2.21
16	VIIIb	right	23.5	-50.5	-58.0	5	0.016	2.21
17	IX	left	-6.5	-62.5	-56.5	4	0.017	2.18
18	Crus II	left	-6.5	-85.0	-43.0	3	0.019	2.14
19	VIIb	left	-32.0	-62.5	-52.0	3	0.019	2.13
20	VIIIb	right	26.5	-40.0	-52.0	1	0.02	2.12
21	Crus II	left	-8.0	-91.0	-35.5	2	0.02	2.11
22	I-IV	right	11.5	-40.0	-10.0	1	0.021	2.08
23	Crus I	right	50.5	-62.5	-32.5	3	0.022	2.07
24	VIIIa	left	-32.0	-43.0	-52.0	1	0.023	2.05

25	I-IV	right	11.5	-38.5	-13.0	1	0.024	2.04
26	VI	right	37.0	-58.0	-23.5	2	0.024	2.03
27	X	left	-24.5	-34.0	-41.5	1	0.024	2.03
28	IX	left	-6.5	-46.0	-50.5	1	0.024	2.02
Figure 8A: No US post CS+ x prediction error PPI negative connectivity with VTA during reextinction, t-test, p < 0.05, uncorrected								
1	I-IV	right	4.0	-50.5	-14.5	38	0.001	3.33
2	VIIIb	left	-9.5	-43.0	-58.0	18	0.001	3.20
3	white matter		-12.5	-47.5	-26.5	31	0.002	3.03
4	Extended cluster	left V (20), white matter (9), left I-IV (9)						
	white matter		-18.5	-31.0	-31.0	38	0.003	2.93
	V	left	-26.0	-35.5	-28.0		0.016	2.22
5	V	right	31.0	-38.5	-25.0	24	0.003	2.86
6	VI	left	-35.0	-38.5	-32.5	42	0.003	2.86
7	Crus II	right	7.0	-80.5	-41.5	29	0.004	2.79
8	Extended cluster	right IX (116), right VIIIb (90), white matter (22), right VIIIa (17)						
	IX	right	14.5	-55.0	-50.5	245	0.004	2.76
	IX	right	8.5	-56.5	-44.5		0.005	2.69
	VIIIb	right	17.5	-44.5	-50.5		0.016	2.22
9	Extended cluster	white matter (187), left Crus II (25), left Crus I (17), left VIIIa (9), left VI (4), left DN (1)						
	white matter		-30.5	-59.5	-40.0	243	0.005	2.72
	white matter		-24.5	-55.0	-37.0		0.006	2.60
	VIIIa	left	-29.0	-52.0	-46.0		0.014	2.26
10	V	right	32.5	-34.0	-34.0	6	0.005	2.67
11	white matter		-12.5	-40.0	-31.0	92	0.007	2.55
12	I-IV	left	-5.0	-49.0	-13.0	19	0.008	2.53
13	VI	vermal	1.0	-77.5	-25.0	4	0.008	2.51
14	VIIb	right	11.5	-71.5	-49.0	29	0.009	2.48
15	VIIIa	left	-8.0	-70.0	-53.5	24	0.009	2.44
16	white matter		-14.0	-53.5	-40.0	15	0.011	2.38
17	Crus I	right	32.5	-77.5	-37.0	11	0.011	2.36
18	VIIIb	left	-18.5	-44.5	-52.0	11	0.012	2.35
19	Crus II	right	14.5	-85.0	-37.0	7	0.013	2.31
20	Crus I	left	-17.0	-82.0	-25.0	5	0.014	2.26
21	Crus I	left	-48.5	-47.5	-40.0	1	0.014	2.26
22	IX	left	-5.0	-52.0	-41.5	5	0.015	2.24
23	X	vermal	-2.0	-50.5	-32.5	5	0.016	2.22
24	Crus I	right	44.5	-65.5	-43.0	2	0.019	2.14
25	V	left	-23.0	-46.0	-17.5	1	0.02	2.12
26	Crus II	left	-14.0	-85.0	-40.0	6	0.02	2.11

27	IX	left	-5.0	-49.0	-37.0	1	0.021	2.10
28	Crus I	left	-11.0	-71.5	-31.0	4	0.021	2.09
29	white matter		-27.5	-44.5	-35.5	1	0.023	2.06
30	Crus I	right	46.0	-61.0	-43.0	1	0.023	2.04
31	Crus II	right	40.0	-44.5	-44.5	1	0.023	2.04
Figure 8B: No US post CS+ x prediction error PPI positive connectivity with VTA during extinction, recall, reacquisition and reextinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	left Crus I (158), left VI (142), left V (14)						
	VI	left	-32.0	-58.0	-25.0	314	0.001	3.53
	VI	left	-21.5	-52.0	-19.0		0.002	2.99
	Crus I	left	-38.0	-65.5	-32.5		0.003	2.90
2	I-IV	left	-14.0	-37.0	-25.0	105	0.001	3.30
3	Extended cluster	left X (23), left VIIIb (11), left VIIIa (6)						
	X	left	-21.5	-34.0	-46.0	40	0.001	3.25
	VIIIa	left	-29.0	-40.0	-50.5		0.016	2.21
4	Extended cluster	right V (225), right VI (32), white matter (21), right I-IV (16)						
	V	right	17.5	-47.5	-22.0	294	0.001	3.19
	V	right	13.0	-53.5	-19.0		0.003	2.94
5	Extended cluster	left I-IV (33), right I-IV (32)						
	I-IV	left	-5.0	-43.0	-22.0	65	0.002	3.13
	I-IV	right	4.0	-43.0	-20.5		0.004	2.76
6	VIIIb	left	-24.5	-46.0	-58.0	31	0.002	3.08
7	VIIIb	right	19.0	-59.5	-59.5	53	0.002	3.03
8	VIIb	left	-36.5	-41.5	-47.5	24	0.002	3.01
9	VI	left	-23.0	-61.0	-28.0	151	0.003	2.85
10	VIIIb	right	25.0	-52.0	-58.0	13	0.005	2.69
11	Crus II	vermal	-2.0	-79.0	-29.5	16	0.005	2.68
12	white matter		41.5	-53.5	-38.5	29	0.005	2.67
13	VIIIb	right	13.0	-40.0	-52.0	3	0.005	2.67
14	V	right	2.5	-61.0	-23.5	36	0.006	2.65
15	VI	left	-6.5	-73.0	-22.0	25	0.006	2.62
16	VI	right	8.5	-73.0	-22.0	23	0.007	2.59
17	VIIb	left	-26.0	-68.5	-55.0	8	0.007	2.57
18	Extended cluster	right Crus I (33), right VI (27)						
	Crus I	right	32.5	-67.0	-32.5	60	0.007	2.55
	VI	right	22.0	-62.5	-29.5		0.014	2.26
19	VIIIb	left	-15.5	-38.5	-50.5	4	0.009	2.47
20	VI	right	34.0	-56.5	-23.5	22	0.009	2.45
21	V	right	10.0	-61.0	-19.0	11	0.009	2.45
22	Crus II	left	-21.5	-79.0	-52.0	6	0.01	2.41
23	Crus II	left	-29.0	-79.0	-52.0	21	0.01	2.41

24	V	right	2.5	-65.5	-7.0	20	0.01	2.41
25	I-IV	left	-8.0	-43.0	-8.5	12	0.01	2.41
26	Crus I	left	-45.5	-58.0	-32.5	9	0.01	2.41
27	VI	right	28.0	-67.0	-23.5	50	0.013	2.32
28	VIIIa	right	10.0	-70.0	-56.5	3	0.013	2.30
29	VIIIa	left	-8.0	-68.5	-44.5	6	0.014	2.27
30	VIIIb	right	23.5	-46.0	-58.0	3	0.015	2.25
31	VI	left	-9.5	-67.0	-26.5	5	0.015	2.24
32	VI	left	-33.5	-43.0	-25.0	6	0.015	2.23
33	IX	right	2.5	-55.0	-47.5	5	0.017	2.20
34	VIIIa	left	-30.5	-47.5	-52.0	3	0.017	2.18
35	I-IV	right	14.5	-40.0	-28.0	3	0.018	2.15
36	VIIIa	vermal	5.5	-70.0	-44.5	4	0.019	2.15
37	Crus II	left	-33.5	-80.5	-47.5	2	0.022	2.08
38	Crus II	left	-15.5	-80.5	-50.5	1	0.022	2.08
39	IX	right	1.0	-62.5	-55.0	2	0.022	2.07
40	Crus I	left	-32.0	-79.0	-31.0	1	0.022	2.07
41	Crus II	left	-17.0	-85.0	-49.0	1	0.023	2.04
42	Crus II	right	26.5	-80.5	-41.5	1	0.024	2.03
43	X	right	23.5	-34.0	-47.5	1	0.024	2.03
44	Crus II	left	-47.0	-56.5	-46.0	1	0.025	2.02
Figure 8B: No US post CS+ x prediction error PPI negative connectivity with VTA during extinction, recall, reacquisition and reextinction, t-test, p < 0.05, uncorrected								
1	Extended cluster	white matter (515), left V (72), left IX (25), left I-IV (14), left DN (12), left VI (10), vermal X (7), left Crus I (6), left Crus II (4), left X (2)						
	white matter		-18.5	-31.0	-32.5	667	0.001	3.24
	white matter		-24.5	-53.5	-37.0		0.002	3.12
	V	left	-26.0	-37.0	-31.0		0.002	3.00
2	white matter		-9.5	-70.0	-34.0	21	0.004	2.77
3	I-IV	right	4.0	-52.0	-13.0	16	0.006	2.62
4	white matter		10.0	-70.0	-35.5	6	0.006	2.62
5	Extended cluster	white matter (35), right DN (2)						
	white matter		23.5	-52.0	-37.0	37	0.006	2.61
	white matter		16.0	-49.0	-35.5		0.017	2.18
6	white matter		7.0	-46.0	-32.5	10	0.009	2.48
7	VIIIa	right	8.5	-70.0	-50.5	9	0.009	2.47
8	VIIIa	left	-9.5	-68.5	-55.0	4	0.009	2.44
9	Crus II	right	13.0	-82.0	-50.5	2	0.011	2.38
10	Crus II	right	8.5	-80.5	-40.0	9	0.013	2.32
11	Crus II	left	-21.5	-80.5	-43.0	5	0.013	2.30
12	VIIIb	left	-8.0	-46.0	-59.5	12	0.013	2.30

13	white matter		2.5	-59.5	-29.5	5	0.013	2.29
14	Crus I	right	41.5	-46.0	-41.5	5	0.015	2.26
15	white matter		-11.0	-46.0	-25.0	3	0.016	2.21
16	IX	right	5.5	-50.5	-62.5	3	0.017	2.19
17	white matter		34.0	-46.0	-43.0	4	0.018	2.15
18	VIIIb	right	13.0	-61.0	-55.0	1	0.019	2.14
19	VIIIb	right	26.5	-43.0	-44.5	2	0.02	2.12
20	Crus II	left	-12.5	-85.0	-40.0	2	0.021	2.10

Table S12. fMRI psychophysiological interaction (PPI) activation clusters (p < 0.05, uncorrected) related to connectivity with the ventral tegmental area (VTA) during unexpected omission of the unconditioned stimulus (US) using a VTA seed (Figure 8). Clusters were identified in the cerebellar cortex and deep cerebellar nuclei (DCN). Up to three local maxima per cluster are reported, separated by at least 8 mm. Coordinates are given in MNI space (x, y, z). Cluster size is reported as number of voxels (voxel volume = 3.375 mm^3). US: unconditioned stimulus; CS: conditioned stimulus; VTA: ventral tegmental area; DCN: deep cerebellar nuclei; DN: dentate nucleus; IN: interposed nucleus; FN: fastigial nucleus; MNI: Montreal Neurological Institute standard brain; t: t-statistic; p_{unc}: uncorrected p-value.

Data availability

The dataset generated and analysed during the current study has been deposited in the ReSeeD Repository (SFB 1280 Extinction Learning, Ruhr University Bochum). Public release will follow completion of the repository's formal publication procedures in accordance with SFB 1280 data management policies.

Acknowledgements

We would like to thank our technician Beate Brol for the drawing and adjustment of DCN and VTA masks. Additionally, we would like to thank Greta Wippich for providing helpful illustrations. The MAGNETOM Terra 7T MRI system used in the study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant number 432647511. This project has received funding from the DFG (project number 316803389 - SFB1280), the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956414 and an individual scholarship for Patrick Pais Pereira from the Hans-Böckler-Stiftung.

Additional information

Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft (DFG)	SFB1280 - Project number 316803389	Enzo Nio Harald H Quick Giorgi Batsikadze Sen Cheng Patrick Pais Pereira Dagmar Timmann Thomas Michael Ernst Metin Üngör Christian Josef Merz Nicolas Diekmann Mykola Petrenko Alice Doubliez
European Union Horizon 2020	Marie Skłodowska-Curie 956414	Enzo Nio Alice Doubliez Dagmar Timmann
Deutsche Forschungsgemeinschaft (DFG)	432647511	Harald H Quick
Hans Böckler Stiftung (Hans-Böckler-Stiftung)	Individual scholarship for Patrick Pais Pereira	Patrick Pais Pereira

Author ORCID iDs

Enzo Nio: https://orcid.org/0000-0001-8006-8993
Patrick Pais Pereira: https://orcid.org/0009-0004-8200-7931
Nicolas Diekmann: https://orcid.org/0000-0003-3638-7617
Mykola Petrenko: https://orcid.org/0000-0001-5032-1927
Alice Doubliez: https://orcid.org/0000-0002-1736-5458
Thomas M Ernst: https://orcid.org/0000-0002-2170-9241
Giorgi Batsikadze: https://orcid.org/0000-0002-0517-0119
Stefan Maderwald: https://orcid.org/0000-0001-8584-5800
Cornelius Deuschl: https://orcid.org/0000-0001-9262-7289
Metin Üngör: https://orcid.org/0000-0001-5453-2860
Sen Cheng: https://orcid.org/0000-0002-6719-8029
Christian J Merz: https://orcid.org/0000-0001-5679-6595
Harald H Quick: https://orcid.org/0000-0001-7964-0219
Dagmar Timmann: https://orcid.org/0000-0003-1935-416X

References

1. **Grupe D. W.,** Nitschke J. B (2013) Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci* **14**:488-501 https://doi.org/10.1038/nrn3524 | PubMed
2. **Papalini S.,** Beckers T., Vervliet B (2020) Dopamine: from prediction error to psychotherapy. *Transl. Psychiatry* **10**:164 https://doi.org/10.1038/s41398-020-0814-x | PubMed
3. **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 https://doi.org/10.1146/annurev.psych.121208.131631 | PubMed

4. **Lonsdorf T. B.,** et al. (2017) Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci. Biobehav. Rev* **77**:247-285 https://doi.org/10.1016/j.neubiorev.2017.02.026 | PubMed
5. **Luo R.,** et al. (2018) A dopaminergic switch for fear to safety transitions. *Nat. Commun* **9**:2483 https://doi.org/10.1038/s41467-018-04784-7 | PubMed
6. **Salinas-Hernández X. I.,** et al. (2018) Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. *eLife* **7**:e38818 https://doi.org/10.7554/eLife.38818 | PubMed
7. **Kalisch R.,** Gerlicher A. M. V., Duvarci S (2019) A Dopaminergic Basis for Fear Extinction. *Trends Cogn. Sci* **23**:274-277 https://doi.org/10.1016/j.tics.2019.01.013 | PubMed
8. **Craske M. G.,** Liao B., Brown L., Vervliet B (2012) Role of Inhibition in Exposure Therapy. *J. Exp. Psychopathol* **3**:322-345 https://doi.org/10.5127/jep.026511
9. **Schultz W.,** Dickinson A (2000) Neuronal Coding of Prediction Errors. *Annu. Rev. Neurosci* **23**:473-500 https://doi.org/10.1146/annurev.neuro.23.1.473 | PubMed
10. **Salinas-Hernández X. I.,** Zafiri D., Sigurdsson T., Duvarci S (2023) Functional architecture of dopamine neurons driving fear extinction learning. *Neuron* **111**:3854-3870.e5 https://doi.org/10.1016/j.neuron.2023.08.025 | PubMed
11. **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 https://doi.org/10.1038/nature21726 | PubMed
12. **Heffley W.,** Hull C (2019) Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. *eLife* **8**:e46764 https://doi.org/10.7554/eLife.46764 | PubMed
13. **Kostadinov D.,** Beau M., Pozo M. B., Häusser M (2019) Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. *Nat. Neurosci* **22**:950-962 https://doi.org/10.1038/s41593-019-0381-8 | PubMed
14. **Snider R. S.,** Maiti A (1976) Cerebellar contributions to the papez circuit. *J. Neurosci. Res* **2**:133-146 https://doi.org/10.1002/jnr.490020204 | PubMed
15. **Watabe-Uchida M.,** Zhu L., Ogawa S. K., Vamanrao A., Uchida N (2012) Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron* **74**:858-873 https://doi.org/10.1016/j.neuron.2012.03.017 | PubMed
16. **Carta I.,** Chen C. H., Schott A. L., Dorizan S., Khodakhah K (2019) Cerebellar modulation of the reward circuitry and social behavior. *Science* **363** https://doi.org/10.1126/science.aav0581 | PubMed
17. **Pisano T. J.,** et al. (2021) Homologous organization of cerebellar pathways to sensory, motor, and associative forebrain. *Cell Rep* **36**:109721 https://doi.org/10.1016/j.celrep.2021.109721 | PubMed
18. **Ernst T. M.,** et al. (2019) The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. *eLife* **8**:e46831 https://doi.org/10.7554/eLife.46831 | PubMed
19. **Ploghaus A.,** et al. (1999) Dissociating Pain from Its Anticipation in the Human Brain. *Science* **284**:1979-1981 https://doi.org/10.1126/science.284.5422.1979 | PubMed
20. **Bouton M. E** (2002) Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol. Psychiatry* **52**:976-986 https://doi.org/10.1016/s0006-3223(02)01546-9 | PubMed
21. **Doubliez A.,** et al. (2025) Dopaminergic drugs modulate fear extinction-related processes in humans, but effects are mild. *Brain Commun* **7**:fcaf333 https://doi.org/10.1093/braincomms/fcaf333 | PubMed
22. **Diedrichsen J.,** Zotow E (2015) Surface-based display of volume-averaged cerebellar imaging data. *PLoS One* **10**:1-18 https://doi.org/10.1371/journal.pone.0133402 | PubMed
23. **Baek S. J.,** Park J. S., Kim J., Yamamoto Y., Tanaka-Yamamoto K (2022) VTA-projecting cerebellar neurons mediate stress-dependent depression-like behaviors. *eLife* **11**:e72981 https://doi.org/10.7554/eLife.72981 | PubMed

24. **Linnman C.,** et al. (2012) Resting Amygdala and Medial Prefrontal Metabolism Predicts Functional Activation of the Fear Extinction Circuit. *Am. J. Psychiatry* **169**:415-423 https://doi.org/10.1176/appi.ajp.2011.10121780 | PubMed

25. **Kattoor J.,** et al. (2014) Cerebellar contributions to different phases of visceral aversive extinction learning. *Cerebellum* **13**:1-8 https://doi.org/10.1007/s12311-013-0512-9 | PubMed

26. **Faul L.,** et al. (2020) Proximal threats promote enhanced acquisition and persistence of reactive fear-learning circuits. *Proc. Natl. Acad. Sci. U. S. A* **117**:16678-16689 https://doi.org/10.1073/pnas.2004258117 | PubMed

27. **Labrenz F.,** et al. (2022) Temporal dynamics of fMRI signal changes during conditioned interoceptive pain-related fear and safety acquisition and extinction. *Behav. Brain Res* **427** https://doi.org/10.1016/j.bbr.2022.113868 | PubMed

28. **Ploghaus A.,** et al. (2000) Learning about pain: The neural substrate of the prediction error for aversive events. *Proc. Natl. Acad. Sci* **97**:9281-9286 https://doi.org/10.1073/pnas.160266497 | PubMed

29. **Yágüez L.,** et al. (2005) Brain response to visceral aversive conditioning: A functional magnetic resonance imaging study. *Gastroenterology* **128**:1819-1829 https://doi.org/10.1053/j.gastro.2005.02.068 | PubMed

30. **Batsikadze G.,** et al. (2022) The cerebellum contributes to context-effects during fear extinction learning: A 7T fMRI study. *Neuroimage* **253**:119080 https://doi.org/10.1016/j.neuroimage.2022.119080 | PubMed

31. **Doubliez A.,** et al. (2023) The cerebellum and fear extinction: evidence from rodent and human studies. *Front. Syst. Neurosci* **17**:1-8 https://doi.org/10.3389/fnsys.2023.1166166 | PubMed

32. **Fullana M. A.,** et al. (2016) Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Mol. Psychiatry* **21**:500-508 https://doi.org/10.1038/mp.2015.88 | PubMed

33. **Ramnani N.,** Elliott R., Athwal B. S., Passingham R. E (2004) Prediction error for free monetary reward in the human prefrontal cortex. *Neuroimage* **23**:777-786 https://doi.org/10.1016/j.neuroimage.2004.07.028 | PubMed

34. **Kruithof E. S.,** Klaus J., Schutter D. J. L. G (2023) The human cerebellum in reward anticipation and outcome processing: An activation likelihood estimation meta-analysis. *Neurosci. Biobehav. Rev* **149**:105171 https://doi.org/10.1016/j.neubiorev.2023.105171 | PubMed

35. **Nettekoven C.,** et al. (2024) A hierarchical atlas of the human cerebellum for functional precision mapping. *Nat. Commun* **15**:8376 https://doi.org/10.1038/s41467-024-52371-w | PubMed

36. **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 https://doi.org/10.1016/j.neuroimage.2018.01.082 | PubMed

37. **Camilleri J. A.,** et al. (2018) Definition and characterization of an extended multiple-demand network. *Neuroimage* **165**:138-147 https://doi.org/10.1016/j.neuroimage.2017.10.020 | PubMed

38. **Rescorla R. A.,** Wagner A. R (1972) A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In: Black A. H., Prokasy W. F. (Eds). *Classical Conditioning: Current Research and Theory* Appleton-Century-Crofts. pp. 64-99

39. **Pearce J. M.,** Hall G (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev* **87**:532-552 https://doi.org/10.1037/0033-295x.87.6.532 | PubMed

40. **Koenig S.,** Uengoer M., Lachnit H (2017) Attentional bias for uncertain cues of shock in human fear conditioning: Evidence for attentional learning theory. *Front. Hum. Neurosci* **11**:1-13 https://doi.org/10.3389/fnhum.2017.00266 | PubMed

41. **Cai L. X.,** et al. (2020) Distinct signals in medial and lateral VTA dopamine neurons modulate fear extinction at different times. *eLife* **9**:e54936 https://doi.org/10.7554/eLife.54936 | PubMed

42. **Esser R.,** Korn C. W., Ganzer F., Haaker J (2021) L-DOPA modulates activity in the vmPFC, nucleus accumbens, and VTA during threat extinction learning in humans. *eLife* **10**:e65280 https://doi.org/10.7554/eLife.65280 | PubMed

43. **D'Ardenne K.,** McClure S. M., Nystrom L. E., Cohen J. D (2008) BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. *Science* **319**:1264-1267 https://doi.org/10.1126/science.1150605 | PubMed

44. **Schott B. H.,** et al. (2008) Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *J. Neurosci* **28**:14311-14319 https://doi.org/10.1523/jneurosci.2058-08.2008 | PubMed

45. **Kostadinov D.,** Häusser M (2022) Reward signals in the cerebellum: Origins, targets, and functional implications. *Neuron* **110**:1290-1303 https://doi.org/10.1016/j.neuron.2022.02.015 | PubMed

46. **Logothetis N. K.,** Wandell B. A (2004) Interpreting the BOLD signal. *Annu. Rev. Physiol* **66**:735-769 https://doi.org/10.1146/annurev.physiol.66.082602.092845 | PubMed

47. **Logothetis N. K.,** Pauls J., Augath M., Trinath T., Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**:150-157 https://doi.org/10.1038/35084005 | PubMed

48. **Logothetis N. K** (2008) What we can do and what we cannot do with fMRI. *Nature* **453**:869-878 https://doi.org/10.1038/nature06976 | PubMed

49. **Shahshahani L.,** King M., Nettekoven C., Ivry R. B., Diedrichsen J (2024) Selective recruitment of the cerebellum evidenced by task-dependent gating of inputs. *eLife* **13**:e96386 https://doi.org/10.7554/eLife.96386 | PubMed

50. **Gagliano G.,** et al. (2022) Non-Linear Frequency Dependence of Neurovascular Coupling in the Cerebellar Cortex Implies Vasodilation–Vasoconstriction Competition. *Cells* **11** https://doi.org/10.3390/cells11061047 | PubMed

51. **Mapelli L.,** et al. (2017) Granular layer neurons control cerebellar neurovascular coupling through an NMDA receptor/NO-dependent system. *J. Neurosci* **37**:1340-1351 https://doi.org/10.1523/jneurosci.2025-16.2016 | PubMed

52. **Zheng N.,** et al. (2023) Investigations of brain-wide functional and structural networks of dopaminergic and CamKIIα-positive neurons in VTA with DREADD-fMRI and neurotropic virus tracing technologies. *J. Transl. Med* **21**:1-14 https://doi.org/10.1186/s12967-023-04362-6 | PubMed

53. **Düzel E.,** et al. (2009) Functional imaging of the human dopaminergic midbrain. *Trends Neurosci* **32**:321-328 https://doi.org/10.1016/j.tins.2009.02.005 | PubMed

54. **Ikai Y.,** Takada M., Shinonaga Y., Mizuno N (1992) Dopaminergic and non-dopaminergic neurons in the ventral tegmental area of the rat project, respectively, to the cerebellar cortex and deep cerebellar nuclei. *Neuroscience* **51**:719-728 https://doi.org/10.1016/0306-4522(92)90310-x | PubMed

55. **Li C.,** et al. (2023) Purkinje cell dopaminergic inputs to astrocytes regulate cerebellar-dependent behavior. *Nat. Commun* **14** https://doi.org/10.1038/s41467-023-37319-w | PubMed

56. **Guarque-Chabrera J.,** Oñate M., Vera J., Khodakhah K. (2024) Exploring monosynaptic midbrain-to-deep cerebellar nuclei projections. In: Abstracts, Neuroscience 2024.

57. **Jo Y. S.,** Heymann G., Zweifel L. S (2018) Dopamine Neurons Reflect the Uncertainty in Fear Generalization. *Neuron* **100**:916-925.e3 https://doi.org/10.1016/j.neuron.2018.09.028 | PubMed

58. **Moriya S.,** et al. (2018) Acute Aversive Stimuli Rapidly Increase the Activity of Ventral Tegmental Area Dopamine Neurons in Awake Mice. *Neuroscience* **386**:16-23 https://doi.org/10.1016/j.neuroscience.2018.06.027 | PubMed

59. **Rouhani N.,** Niv Y (2021) Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife* **10**:e61077 https://doi.org/10.7554/eLife.61077 | PubMed

60. **Montague P. R.,** Dayan P., Sejnowski T. J (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci* **16**:1936-1947 https://doi.org/10.1523/jneurosci.16-05-01936.1996 | PubMed

61. **Barto A. G.** (2018) Adaptive Critics and the Basal Ganglia. In: Hook J. C., Davis J. L., Beiser D. G. (Eds). *Models of Information Processing in the Basal Ganglia* The MIT Press. pp. 215-232 https://doi.org/10.7551/mitpress/4708.003.0018

62. **Hull C** (2020) Prediction signals in the cerebellum: Beyond supervised motor learning. *eLife* **9**:e54073 https://doi.org/10.7554/eLife.54073 | PubMed

63. **Corlett P. R.,** Mollick J. A., Kober H (2022) Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology* **47**:1339-1349 https://doi.org/10.1038/s41386-021-01264-3 | PubMed

64. **Batsikadze G.,** et al. (2024) Mild Deficits in Fear Learning: Evidence from Humans and Mice with Cerebellar Cortical Degeneration. *eNeuro* **11**:ENEURO.0365-23.2023 https://doi.org/10.1523/eneuro.0365-23.2023 | PubMed

65. **Gainotti G** (2019) Emotions and the Right Hemisphere: Can New Data Clarify Old Models?. *Neuroscientist* **25**:258-270 https://doi.org/10.1177/1073858418785342 | PubMed

66. **Güntürkün O.,** Ströckens F., Ocklenburg S (2020) Brain lateralization: A comparative perspective. *Physiol. Rev* **100**:1019-1063 https://doi.org/10.1152/physrev.00006.2019 | PubMed

67. **Pakusch J.,** et al. (2025) Evidence for lateralization of fear emotions in the cerebellum. *J. Neurol* **272**:1-5 https://doi.org/10.1007/s00415-025-13183-0 | PubMed

68. **Liu G.,** et al. (2018) Praising others differently: Neuroanatomical correlates to individual differences in trait gratitude and elevation. *Soc. Cogn. Affect. Neurosci* **13**:1225-1234 https://doi.org/10.1093/scan/nsy093 | PubMed

69. **Zhong S.,** Su T., Gong J., Huang L., Wang Y (2023) Brain functional alterations in patients with anorexia nervosa: A meta-analysis of task-based functional MRI studies. *Psychiatry Res* **327**:115358 https://doi.org/10.1016/j.psychres.2023.115358 | PubMed

70. **E K,** Chen S. A., Ho M. R., Desmond J. E (2014) A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Hum. Brain Mapp* **35**:593-615 https://doi.org/10.1002/hbm.22194 | PubMed

71. **Stoodley C. J.,** Schmahmann J. D (2009) Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *Neuroimage* **44**:489-501 https://doi.org/10.1016/j.neuroimage.2008.08.039 | PubMed

72. **Pierce J. E.,** Thomasson M., Voruz P., Selosse G., Péron J (2023) Explicit and Implicit Emotion Processing in the Cerebellum: A Meta-analysis and Systematic Review. *Cerebellum* **22**:852-864 https://doi.org/10.1007/s12311-022-01459-4 | PubMed

73. **Bartolomeo P.,** Seidel Malkinson T (2019) Hemispheric lateralization of attention processes in the human brain. *Curr. Opin. Psychol* **29**:90-96 https://doi.org/10.1016/j.copsyc.2018.12.023 | PubMed

74. **Friston K. .,** et al. (1997) Psychophysiological and Modulatory Interactions in Neuroimaging. *Neuroimage* **6**:218-229 https://doi.org/10.1006/nimg.1997.0291 | PubMed

75. **O'Reilly J. X.,** Woolrich M. W., Behrens T. E. J., Smith S. M., Johansen-Berg H (2012) Tools of the trade: Psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci* **7**:604-609 https://doi.org/10.1093/scan/nss055 | PubMed

76. **Diedrichsen J.,** et al. (2011) Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *Neuroimage* **54**:1786-1794 https://doi.org/10.1016/j.neuroimage.2010.10.035 | PubMed

77. **de Hollander G.,** Keuken M. C., van der Zwaag W., Forstmann B. U., Trampel R (2017) Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7 T and 3 T. *Hum. Brain Mapp* **38**:3226-3248 https://doi.org/10.1002/hbm.23586 | PubMed

78. **Pauli W. M.,** Nili A. N., Tyszka J. M (2018) A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* **5**:180063 https://doi.org/10.1038/sdata.2018.63 | PubMed

79. **Trutti A. C.,** et al. (2021) A probabilistic atlas of the human ventral tegmental area (VTA) based on 7 Tesla MRI data. *Brain Struct. Funct* **226**:1155-1167 https://doi.org/10.1007/s00429-021-02231-w | PubMed

80. **Nio E.,** et al. (2023) 7T fMRI study on extinction learning in young and healthy participants. *Cent. Open Sci* https://doi.org/10.17605/OSF.IO/2PXWE

81. **Mumford J. A** (2012) A power calculation guide for FMRI studies. *Soc. Cogn. Affect. Neurosci* **7**:738-742 https://doi.org/10.1093/scan/nss059 | PubMed

82. **Henry J. D.,** Crawford J. R (2005) The short-form version of the Depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *Br. J. Clin. Psychol* **44**:227-239 https://doi.org/10.1348/014466505x29657 | PubMed

83. **Oldfield R. C** (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**:97-113 https://doi.org/10.1016/0028-3932(71)90067-4 | PubMed

84. **Inoue L.,** et al. (2020) Interaction of fear conditioning with eyeblink conditioning supports the sensory gating hypothesis of the amygdala in men. *eNeuro* **7**:1-15 https://doi.org/10.1523/eneuro.0128-20.2020 | PubMed

85. **Otto T.,** Wolf O. T., Merz C. J (2023) EDA-Analysis App (5.11). *Zenodo* https://doi.org/10.5281/zenodo.7965376

86. **Boucsein W.** (2012) *Electrodermal Activity* Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-1126-0

87. **Kret M. E.,** Sjak-Shie E. E (2019) Preprocessing pupil size data: Guidelines and code. *Behav. Res. Methods* **51**:1336-1342 https://doi.org/10.3758/s13428-018-1075-y | PubMed

88. **Jentsch V. L.,** Wolf O. T., Merz C. J (2020) Temporal dynamics of conditioned skin conductance and pupillary responses during fear acquisition and extinction. *Int. J. Psychophysiol* **147**:93-99 https://doi.org/10.1016/j.ijpsycho.2019.11.006 | PubMed

89. **Mathôt S.,** Fabius J., van Heusden E., Van der Stigchel S (2018) Safe and sensible baseline correction of pupil-size data. *Behav. Res. Methods* 94-106 https://doi.org/10.3758/s13428-017-1007-2 | PubMed

90. **Brunner E.,** Domhof S., Langer F (2002) *Nonparametric Analysis of Longitudinal Data in Factorial Experiments* New York, NY: John Wiley & Sons.

91. **Shah D. a,** Madden L. V. (2004) Nonparametric analysis of ordinal data. *Phytopathology* **94**:33-43 https://doi.org/10.1094/phyto.2004.94.1.33 | PubMed

92. **Noguchi K.,** Gel Y. R., Brunner E., Konietschke F (2012) nparLD : An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *J. Stat. Softw* **50** https://doi.org/10.18637/jss.v050.i12

93. **Bathke A. C.,** Schabenberger O., Tobias R. D., Madden L. V (2009) Greenhouse-Geisser adjustment and the ANOVA-type statistic: Cousins or twins?. *Am. Stat* **63**:239-246 https://doi.org/10.1198/tast.2009.08187

94. **Sutton R. S.,** Barto A. G (2018) *Reinforcement Learning: An Introduction* MA: MIT Press, Cambridge.

95. **LeCun Y.,** Bengio Y., Hinton G. (2015) Deep learning. *Nature* **521**:436-444 https://doi.org/10.1038/nature14539 | PubMed

96. **Lin L.-J. J** Self-improvement Based On Reinforcement Learning, Planning and Teaching. In: Proceedings of the 8th International Workshop on Machine Learning, ICML 1991. pp. 323-327 https://doi.org/10.1016/b978-1-55860-200-7.50067-2

97. **Rumelhart D. E.,** Hinton G. E., Williams R. J (1986) Learning representations by back-propagating errors. *Nature* **323**:533-536 https://doi.org/10.1038/323533a0

98. **Schaul T.,** Quan J., Antonoglou I., Silver D. (2016) Prioritized Experience Replay. *arXiv* https://doi.org/10.48550/arXiv.1511.05952

99. **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 https://doi.org/10.1002/mrm.23108 | PubMed

100. **Stirnberg R.,** Stöcker T (2021) Segmented K-space blipped-controlled aliasing in parallel imaging for high spatiotemporal resolution EPI. *Magn. Reson. Med* **85**:1540-1551 https://doi.org/10.1002/mrm.28486 | PubMed

101. **Eckstein K.,** et al. (2018) Computationally Efficient Combination of Multi-channel Phase Data From Multi-echo Acquisitions (ASPIRE). *Magn. Reson. Med* **79**:2996-3006 https://doi.org/10.1002/mrm.26963 | PubMed

102. **Eckstein K.,** et al. (2021) Improved susceptibility weighted imaging at ultra-high field using bipolar multi-echo acquisition and optimized image processing: CLEAR-SWI. *Neuroimage* **237**:118175 https://doi.org/10.1016/j.neuroimage.2021.118175 | PubMed

103. **Gallichan D.,** Marques J. P (2017) Optimizing the acceleration and resolution of three-dimensional fat image navigators for high-resolution motion correction at 7T. *Magn. Reson. Med* **77**:547-558 https://doi.org/10.1002/mrm.26127 | PubMed

104. **Marques J. P.,** et al. (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage* **49**:1271-1281 https://doi.org/10.1016/j.neuroimage.2009.10.002 | PubMed

105. **Gallichan D.,** Marques J. P., Gruetter R (2016) Retrospective correction of involuntary microscopic head movement using highly accelerated fat image navigators (3D FatNavs) at 7T. *Magn. Reson. Med* **75**:1030-1039 https://doi.org/10.1002/mrm.25670 | PubMed

106. **Gaser C.,** et al. (2024) CAT: a computational anatomy toolbox for the analysis of structural MRI data. *Gigascience* **13** https://doi.org/10.1093/gigascience/giae049 | PubMed

107. **Jenkinson M.,** Pechaud M., Smith M (2005) BET2: MR-based estimation of brain, skull and scalp surfaces. In: Eleventh Annual Meeting of the Organization for Human Brain Mapping.

108. **Tustison N. J.,** et al. (2014) Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage* **99**:166-179 https://doi.org/10.1016/j.neuroimage.2014.05.044 | PubMed

109. **Rorden C** (2025) MRIcroGL: voxel-based visualization for neuroimaging. *Nat. Methods* **22**:1613-1614 https://doi.org/10.1038/s41592-025-02763-7 | PubMed

110. **Diedrichsen J** (2006) A spatially unbiased atlas template of the human cerebellum. *Neuroimage* **33**:127-138 https://doi.org/10.1016/j.neuroimage.2006.05.056 | PubMed

111. **Yushkevich P. A.,** et al. (2006) User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **31**:1116-1128 https://doi.org/10.1016/j.neuroimage.2006.01.015 | PubMed

112. **McLaren D. G.,** Ries M. L., Xu G., Johnson S. C (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage* **61**:1277-1286 https://doi.org/10.1016/j.neuroimage.2012.03.068 | PubMed

113. **Friston K. J.,** Harrison L., Penny W (2003) Dynamic causal modelling. *Neuroimage* **19**:1273-1302 https://doi.org/10.1016/s1053-8119(03)00202-7 | PubMed

114. **Stephan K. E.,** et al. (2010) Ten simple rules for dynamic causal modeling. *Neuroimage* **49**:3099-3109 https://doi.org/10.1016/j.neuroimage.2009.11.015 | PubMed

115. **Zeidman P.,** et al. (2019) A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *Neuroimage* **200**:174-190 https://doi.org/10.1016/j.neuroimage.2019.06.031 | PubMed

116. **Zeidman P.,** et al. (2019) A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *Neuroimage* **200**:12-25 https://doi.org/10.1016/j.neuroimage.2019.06.032 | PubMed

Peer reviews

Reviewer #1 (Public review):

Nio and colleagues address an important question about how the cerebellum and ventral tegmental area (VTA) contribute to extinction learning of conditioned fear associations. This work tackles a critical gap in the existing literature and provides new insights into this question in humans through the use of high-field neuroimaging with robust methodology. The presented results are novel and will broadly interest both the extinction learning and cerebellar research communities. As such, this is a very timely and important contribution.

Strengths:

The core finding - coupling of cerebellum and VTA as a reward-like prediction errors during fear extinction - is novel and addresses a genuine gap in the literature. Also the paradigm spanning several sessions, a well-powered sample, 7T imaging and complementary analytical approaches to target the question is commendable.

Weaknesses:

The authors have satisfactorily addressed the concerns raised in the previous version of the manuscript. Several results, as well as conclusions drawn from them, still rest on trend-level evidence, although the revised presentation of the results now provides a more balanced interpretation of these findings.

https://doi.org/10.7554/eLife.105399.2.sa1

Author Response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Nio and colleagues address an important question about how the cerebellum and ventral tegmental area (VTA) contribute to the extinction learning of conditioned fear associations. This work tackles a critical gap in the existing literature and provides new insights into this question in humans through the use of high-field neuroimaging with robust methodology. The presented results are novel and will broadly interest both the extinction learning and cerebellar research communities. As such, this is a very timely and impactful manuscript. However, there are several points that could be addressed during the review process to strengthen the claims and enhance their value for readers and the broader scientific community.*
>
> *(1) Reward Interpretation and Skin Conductance Responses (SCR)*
>
> *A central premise of the manuscript is that 'unexpected omissions of expected aversive events' are rewarding, which plays a critical role in extinction learning. The authors also suggest that the cerebellum is involved in reward processing. However, it is unclear how this conclusion can be directly drawn from their task, which does not explicitly model 'reward.' Instead, the interpretation relies on SCR, which seems more indicative of association or prediction rather than reward per se. Is SCR a valid metric of reward experienced during the extinction of feared associations? Or could these findings reflect processes tied more closely to predictive learning? Please, discuss.*

We thank the reviewer for raising this important point. We agree that skin conductance responses (SCRs) do not directly index reward. More generally, SCRs reflect autonomic arousal in response to salient or motivationally significant stimuli and are closely linked to expectancy and contingency awareness. In our study, SCRs served as a read-out of the participants' expectation of a US, and were used to fit the hyperparameters of a reinforcement-learning-based deep learning model, which then provided per-trial estimates of prediction and prediction error values. These estimates capture predictive learning about the occurrence of the aversive US, rather than reward per se. The interpretation of unexpected US omissions as "reward-like" prediction errors relies on prior literature, particularly rodent studies showing that dopaminergic neurons in the VTA respond to omitted aversive stimuli and drive extinction learning via projections to the nucleus accumbens (Kalisch et al., 2019; Salinas-Hernández et al., 2018, 2023). We therefore interpret our cerebellar activations during unexpected omissions as being compatible with the processing of reward-like prediction errors, while acknowledging that this inference is indirect.

To clarify this reasoning, we made revisions to the Introduction and Discussion to (i) state explicitly that SCRs do not directly measure reward but were incorporated into the reinforcement learning model as an index of autonomic arousal related to US expectancy and predictive learning, and (ii) consistently replace the term "reward prediction error" with "reward-like prediction error" throughout.

> *(2) Reinforcement Agent and SCR Modeling*
>
> *The modeling approach with the deep reinforcement agent treats SCR as a personalized expectation of shock for a given trial. However, this interpretation seems misaligned with participants' actual experience - they are aware of the shock but exhibit evolving responses to it over time. Why is this operationalization useful or valid? It would benefit the manuscript to provide a clearer justification for this approach.*

This point is well taken. We did not collect trial-by-trial expectancy ratings, as frequent button-box responses would have induced cerebellar activations unrelated to fear (extinction) learning. Subjective expectancy was assessed only at the end of each experimental phase. As frequently done in the human fear conditioning literature, we used trial-by-trial SCR data (Lonsdorf et al., 2017). Although SCRs show correspondence with US expectancy ratings, they are inherently noisy and show substantial variability across trials and participants (Constantinou et al., 2021). Therefore, individual trial-by-trial responses cannot be used to directly infer US predictions. Accordingly, we used group-averaged SCR data to fit model hyperparameters in a grid search across parameter settings. The best-fitting hyperparameters were then applied to 100 randomly initialized agents, and their outputs were averaged to generate trial-wise estimates of predictions and prediction errors. These averaged values were used as parametric modulators in the fMRI analyses. We have revised the Introduction and Methods to make this procedure clearer.

> *(3) Clarity and Visualization of Results*
>
> *The results section is challenging to follow, and the visualization and quantification of findings could be significantly improved. Terms like 'trending' appear frequently - what does this mean, and is it worth reporting? Adding clear statistical quantifications alongside additional visualizations (e.g., bar or violin plots of group means within specific subregions within the cerebellum, or grouped mean activity in VTA and DCN) would enhance clarity and allow readers to better assess the distribution and systematicity of effects. Furthermore, the figures are overly complex and difficult to read due to the heavy use of abbreviations. Consider splitting figures by either phase of the*

> *experiment or regions, and move some details to the supplemental material for improved readability.*

We agree with the reviewer that the clarity of results can be improved and have revised the manuscript accordingly. Specifically:

(1) We use "trend-level" to refer to uncorrected voxelwise t-maps at $p < 0.05$, and "significant" to refer to TFCE/FWE-corrected effects at $p < 0.05$. This distinction was not sufficiently clear in the original figures. To address this, uncorrected t-maps are now displayed with a grey striped background frame, and colorbar labels have been enlarged to emphasize whether TFCE/FWE-corrected or uncorrected t-values are shown.

(2) We added a supplementary table (Table S7) reporting group-level summary statistics for all fMRI contrasts presented in the manuscript, including group means, standard deviations, effect sizes (Cohen's d), and 95% confidence intervals for cerebellar cortex, cerebellar nuclei, and VTA VOIs. We hope that this helps with the interpretation of effect magnitude and variability across fMRI analyses.

(3) To improve readability, we split overly complex figures: Figure 2 now separates CS-related prediction from US-related presentation contrasts (which are now revised Figures 4 and 5), and Figure 3 separates event-based and parametric modulation contrasts (which are now revised Figures 6 and 7).

(4) We also reduced abbreviations in the figures, and provide full definitions and explanations also including the original abbreviations in the main text and figure captions for clarity.

We considered the suggestion to split figures further by region or by phase. However, we believe it is more informative to present the cerebellar cortex, nuclei, and VTA together for each contrast, and to keep all phases side by side, as this allows readers to directly assess commonalities across phases. We therefore chose to keep the same overall structure, but simplified the figures in other ways (e.g. splitting by contrast type) to improve overall readability. We hope that these changes address the reviewer's concerns by simplifying the presentation, removing abbreviations, and providing clearer quantification of results.

> *(4) Theoretical Context for Paradigm Phases*
>
> *The manuscript benefits from the comprehensive experimental paradigm, which includes multiple phases (acquisition, extinction, recall, reacquisition, re-extinction). This design has great potential for providing a more holistic view of conditioned fear learning and extinction. However, the manuscript lacks clarity on what insights can be drawn from these distinct phases. What theoretical framework underpins the different stages, and how should the results be interpreted in this context? At present, the findings seem like a display of similar patterns across phases without sufficient interpretation. Providing a stronger theoretical rationale and reorganizing the results by experimental phase could significantly improve readability and impact.*

We thank the reviewer for this constructive suggestion. We would first like to mention that the primary aim of this manuscript is not to analyze differences between phases, but rather to highlight the commonalities. Across different learning contexts, we consistently observed reward-like prediction error-related activations in the cerebellum and VTA. This consistency and connectivity between the cerebellum and VTA, despite phase-to-phase differences, is the most important finding of our study.

We agree, however, that the manuscript did not sufficiently explain how each phase differs conceptually, which is important for readers to understand why the consistency of responses

is notable. We therefore expanded the Introduction and Discussion to provide clearer theoretical context for each phase. More specifically, the phases can be understood as follows:

Extinction (day 2): Because acquisition was conducted with a 100% reinforcement rate, unexpected US omissions during initial extinction trials maximize reward-like prediction errors and yield stronger, more uniform expectations across participants compared to a partial reinforcement rate. This phase should therefore provide the clearest opportunity to observe cerebellar-VTA contributions to the processing of reward-like prediction errors.

Recall (day 3): Despite allowing for the consolidation of extinction learning, the recall test often still elicits conditioned fear responses to the CS+, that is, shows spontaneous recovery of the initial fear association (Bouton, 2002). In these trials, the non-occurrence of the US is unexpected. In this context, US omission-related activations reflect reward-like prediction errors during renewed fear responding in the presence of both a fear memory and an extinction memory. This contrasts with extinction training on day 2, where prediction errors arose primarily against the background of the recently acquired fear memory, without a competing extinction memory.

Reacquisition (day 3): Unlike acquisition, reacquisition used a partial reinforcement rate, such that non-reinforced CS+ trials were interspersed between reinforced CS+ trials (similar to the partially reinforced phase used by Ernst et al., 2019). Because reacquisition occurs in the presence of savings, that is, the presence of a previously acquired fear memory, US expectancy increases rapidly following reinforced trials and relearning occurs faster (Bouton, 2004). Importantly, partial reinforcement maintains high US expectancy and therefore allows prediction errors to remain sustained across omission trials (Figure 9).

Reextinction (day 3): Reextinction is an additional extinction phase but without a consolidation interval, and with an already established fear extinction memory. Because reextinction followed the partially reinforced reacquisition phase, prediction errors during early reextinction decayed more slowly than during extinction on day 2 (following the fully reinforced acquisition phase on day 1) (Figure 9). Together, reacquisition and reextinction were designed to maximize the number and persistence of unexpected US omissions, thereby providing additional opportunities to examine reward-like prediction-error signaling.

By clarifying this framework, we aim to show that while the learning context and history differ across phases, the consistent cerebellum-VTA activation and connectivity related to unexpected US omissions underlines the robustness of the effect. We chose not to reorganize the Results by phase, as our central conclusion rests on similarities rather than differences. Instead, we have clarified the theoretical background in the revised manuscript to help readers interpret both the commonalities and the potential sources of variability.

> *(5) Cerebellum-VTA Connectivity Analysis*
>
> *The authors argue that the cerebellum modulates VTA activity, yet they perform the PPI analysis in the reverse direction. Why does this make sense? In their DCM analysis, they found a bidirectional relationship (both cerebellum - VTA and VTA-cerebellum), yet the discussion focused on connectivity from the cerebellum to VTA. A more careful interpretation of the connectivity findings would be useful - especially the strong claims in the discussion on the cerebellum providing the reward signal to the VTA should be tempered.*

We thank the reviewer for highlighting this issue. In our primary analysis, we used the VTA as the PPI seed and observed trend-level connectivity with the cerebellum. When we

d the analysis and used the cerebellar volume of interest (VOI) from the conjunction analysis as the seed, effects in the VTA were substantially weaker. We believe this reflects the broad connectivity profile of the cerebellar VOI (i.e., not specific to the VTA) as well as general limitations of PPI in our study, including the small number of unexpected omission trials and the lack of specificity to reward-like prediction errors (e.g., connectivity also appeared during US presentation). For transparency, we now report the cerebellar-seed PPI results in the Supplementary information (Figure S3). Given their limited robustness, we chose not to include the corresponding VTA maps in the main figures.

Finally, we agree that our conclusions regarding cerebellum-VTA interactions should be framed more cautiously. While the DCM analyses support bidirectional connectivity, our original discussion placed disproportionate emphasis on cerebellum-to-VTA influences. We have revised the text to provide a more balanced interpretation that also considers VTA-to-cerebellum connectivity.

> ***Reviewer #2 (Public review):***
>
> *Summary*
>
> *Building upon the group's previous work, this study used a 3-day threat acquisition, extinction, recall, reextinction, and reacquisition paradigm with 7T imaging to probe the mechanism by which the cerebellum contributes to fear extinction learning. The authors hypothesize this may be via its connection to the VTA, a known modulator of fear extinction due to its role in reward processing. Using complementary analysis methods, the authors demonstrate that activity with the cerebellum, DNC, and VTA is modulated by predictions about the occurrence of the US, which shows regional specificity. They show trend-level evidence that there is increased functional connectivity between the cerebellum and VTA during all phases of the paradigm with unexpected omissions. They also present a DCM which indicates that the cerebellum could positively modulate VTA activity during extinction learning. This study adds to a growing literature supporting the role of the historically overlooked cerebellum in the control of emotions and suggests that an interaction between the cerebellum and VTA should be considered in the existing model of the fear extinction network.*
>
> *Strengths*
>
> *The authors address their research question using a number of complementary methods, including parametric modulation by model-derived expectation parameters, PPI, and DCM, in a logical and easily understood way. I feel the authors provide a balanced interpretation of their findings, presenting numerous interpretations and offering insight with regard to reward vs attention or unsigned prediction errors and the directionality of the interaction they identify. The manuscript is a timely addition to growing literature highlighting the role of the cerebellum in fear conditioning, and emotion generation and regulation more generally.*
>
> *Weaknesses*

> *Subjective and skin conductance responses do not completely support the success of the learning paradigm. For example, CS+/CS- differentiation in both domains persisted after extinction training. I do not feel that this negates the findings of this manuscript, though it raises questions about the parametric modulators used, and the interpretation of the neural mechanisms proposed if they do not strongly relate to updated subjective appraisals (the goal of extinction therapy). My interpretation of the manuscript suggests there are some key results based upon contrasts that have as few as three events; I am a little unsure about the power and reliability of these effects, though I await author clarification on this matter. There are a number of unaddressed deviations from the pre-registered protocol that I have asked the authors to elaborate upon.*

We thank the reviewer for the thoughtful and constructive evaluation of our work. We appreciate that the manuscript and methods were found to be clearly presented, and we welcome the suggestions for clarification and improvement. Below we address the specific concerns regarding extinction learning in behavioral measures, the reliability of event-based contrasts with few trials, and deviations from the preregistration.

Extinction in self-reports and skin conductance responses (SCRs)

The reviewer is correct that CS+/CS- differentiation persisted after extinction. Although there was no differentiation in SCRs at the end of extinction, post-extinction self-reports continued to do so, albeit to a lesser degree, which is in line with previous literature on dissociation of outcome measures during fear conditioning (Lipp et al., 2003). This residual subjective differentiation is also consistent with extinction forming an inhibitory memory trace that suppresses, rather than erases, the original fear association (Bouton, 2002; Milad & Quirk, 2012), and a single extinction session is often insufficient to eliminate differential responding (Craske et al., 2014; Vervliet et al., 2013). However, both measures showed significant effects of extinction learning.

We included additional analyses of self-reports across phases. Importantly, CS+ ratings were significantly reduced during extinction and recall compared to acquisition (all $p \leq 0.001$), whereas CS- ratings remained unchanged (all $p > 0.532$). This pattern demonstrates that the magnitude of the CS+/CS- difference was significantly reduced relative to acquisition, indicating that extinction learning did occur (Doubliez et al., 2025).

For physiological responses, extinction learning was shown in PSRs but not conclusively in SCRs. PSRs showed a significant reduction of CS+ responses across extinction, while CS- responses remained unchanged. SCRs showed a reduction of CS+/CS- differentiation across extinction; however, this effect remained at trend level, as the Stimulus x Time interaction did not reach significance ($p = 0.053$). This pattern is consistent with early differentiation followed by rapid attenuation under the full reinforcement structure of the paradigm (100% reinforcement during acquisition and 0% during extinction). Under such conditions, participants rapidly learn that the US is no longer delivered during extinction, such that physiological responses are largely confined to the first few trials, leaving limited power to detect extinction effects in noisier measures such as SCRs. To address the lower robustness of SCR effects, as recommended by the reviewer, we therefore included PSRs in the main Results section, which provide converging physiological evidence for extinction learning.

Of note, on day 3, both physiological measures and self-reports again showed CS+/CS- differentiation, consistent with spontaneous recovery, a well-established phenomenon reflecting the persistence of the original fear trace after consolidation (Bouton, 2002; Vervliet et al., 2013).

Taken together, these findings demonstrate that the paradigm successfully induced both acquisition and extinction of conditioned fear, even though residual fear responses persisted.

Reliability of event-based contrasts with three trials

The initial decision to use three events for event-based contrasts was based on SCR and PSR data, which showed that differentiation between CS+ and CS- occurred almost exclusively in the first few trials of extinction and recall. Consistent with the full reinforcement described above, prediction errors were expected to be high in the very first extinction trials, and to decay rapidly. Thus, the usual half-block division (e.g., first eight trials) would have included many trials without meaningful prediction errors.

We acknowledge that contrasts based on three trials provide limited statistical power. To address this concern, we added a supplementary table showing summary statistics for contrast estimates in the cerebellar cortex, cerebellar nuclei, and VTA VOIs across all fMRI analyses (Table S7), including both the event-based and parametric modulation approaches. Importantly, the event-based contrasts showed moderate to strong effects despite being restricted to the first three unexpected omission trials. Moreover, the parametric modulation analyses, which incorporate all available trials, yielded results that were consistent with the three-trial event-based contrasts and with the patterns shown in the main figures. This convergence between event-based and parametric approaches strengthens our confidence that the observed effects are reliable.

Deviations from preregistration

We acknowledge that deviations from the preregistered protocol were not fully documented and have now added this information. The main deviation concerned our event-based analyses: while the preregistration planned early vs. late block comparisons, in practice the rapid decay of SCRs under our 100% and 0% reinforcement rates rendered later trials uninformative for prediction error analyses. We therefore focused on the first three trials, when prediction errors are expected to be present. These behavioral findings are also consistent with Doubliez et al. (2025), who used the same paradigm and observed similar rapid SCR decay. Other deviations, such as not reporting exploratory whole-brain DCM analyses, are now clearly stated for transparency.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations for the authors):***
>
> *Minor Point - Paradigm Details*
>
> *Providing additional details about the experimental paradigm in the main text (e.g., the nature of the visual stimuli associated with shocks) would enhance the manuscript's clarity. Some of the information currently in supplementary Figure 5 could be incorporated into the main text to enhance the understanding of the paradigm*

We agree that the current structure reduces clarity, as the paradigm is only explained in detail after the results. To improve readability, we have moved parts of Figure 5 (illustrating the paradigm and scanner setup) to the beginning of the manuscript (now revised Figure 1). In addition, information from Figure 5, including details of the visual stimuli, is now added to the Introduction.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *Methods*
>
> *Can the authors please clarify what part of the task went into [US post CS+ > no US post CS-] contrast? Is this the time immediately after the CS presentations, when the US has just occurred/not occurred, or rather more like the CS+>CS- contrast except including trials confounded by the US (i.e. [CS+/US > CS -])?*

The contrasts are based on an event-related separation of CS and US. The CS was presented for 6 seconds, with its onset modeled in the GLM as a zero-duration event (delta function). The CS offset coincided with either the delivery or omission of the US, which was likewise modeled as a zero-duration event. Thus, CS onset and offset were modeled separately. The no-US events were further distinguished by whether they followed a CS+ or a CS-. Accordingly, we analyzed both CS and US-related contrasts; for example, the CS+ > CS- contrast reflects CS-related differentiation at CS onset (0 s), whereas [US post CS+ > no US post CS-] reflects (no-)US-related activity at CS offset (6 s; US delivered from 5.9-6.0 s). We have added further clarification to the Methods section.

> *I was a bit unclear on what this sentence of the methods meant "Notably, all single trials comprised CS+ trials, with CS- trials also being modeled as single trials to facilitate paired analysis", does this mean that some contrasts had 6 events in total - e.g. the first 3 unexpected omissions vs 3 x CS-. If so, which CS- were selected for the comparison?*

We agree that this sentence was unclear and have revised it. Our intention was to describe that when CS+ trials were modeled as single trials in the GLM (e.g., each CS+ onset and its associated [no-]US event modeled as separate regressors), the CS- trials were modeled in the same way. This ensured that paired analyses would be possible if required.

For reacquisition and reextinction, single-trial modeling was necessary, as the last unexpected omission of reacquisition is also the first unexpected omission of reextinction. Modeling trials separately allows us to examine the first three unexpected US omissions in each phase independently.

The event-based contrasts for unexpected US omissions were defined in line with a previous study of our group. For example, during extinction we contrasted the first three unexpected US omissions following CS+ with all expected omissions following CS- (i.e. [first 3 no US post CS+ > no US post CS-], corresponding to 3 vs. 16 events). The weights of events were automatically scaled by SPM12 so that both sides of the contrast carried equal total weight (e.g. positive events weighted 1/3, negative events weighted -1/16). This procedure matches the approach in Ernst et al. (2019), where in partially reinforced acquisition 6 unexpected omissions after CS+ were contrasted with 16 expected omissions after CS-.

> *More generally, can the authors please comment on the power and reliability of analyses that include only 3 events in a condition [e.g. the first 3 unexpected omissions]?*
>
> *It is not clear if the (US post CS+ > no US post CS-) phases were included. In your pre-registration you say "we will use a "no US post CS+ > no US post CS-" fMRI contrast, where "no US post CS+" designates unexpected omission events in early extinction, early recall (depending on behavioral data which might indicate a return of fear) and a volatile phase (where unexpected omissions occur in the first part of the volatile phase, i.e. reacquisition).", but my reading of the manuscript was that it included both early and late "see 1st level analysis = US post CS+, no US post CS+, no US post CS- separately for each phase; 2nd level = contrast included unexpected omission of the US (no US post CS+ > no US post CS-)". Please clarify and if necessary explain the deviation from preregistration.*

We agree that this point requires clarification. In the preregistration, we planned to divide phases into early and late blocks (no US post CS+ > no US post CS-). However, as already outlined in our response (Reviewer 2, public review response: Reliability of event-based contrasts with three trials), both our preliminary behavioral data and subsequent modeling analyses indicated that differentiation between CS+ and CS- declined extremely rapidly under the 100% reinforcement schedule, leaving likely little or no prediction error beyond the first few trials. Based on this, we adapted the event-based analyses to focus on the first three

unexpected omission trials in extinction, recall, and reextinction, where prediction errors are expected to be present. In reacquisition, only three omission events occur by design (83% reinforcement), so this naturally constrained the analysis to three trials. We now explicitly describe this deviation from the preregistration in the revised manuscript.

As outlined in the same response, we recognize that contrasts based on three trials provide limited statistical power, and addressed this point by providing additional summary VOI statistics of contrast estimates for both event-based and parametric modulation contrasts, which show moderate-to-strong effect sizes and convergence across methods, which we argue supports that using the first three trials is a reliable approach (Reviewer 1, public review response, point (3) Clarity and Visualization of Results).

Finally, with regard to the reviewer's specific question: yes, US post CS+ > no US post CS- contrasts were examined for acquisition training, primarily to demonstrate US-related activation (see revised Figure 3).

> *Results*
>
> *Page 5 + 6: Including the interaction effects for pupil size responses during extinction and reextinction in the SCR section seems unjustified. I appreciate that the SCR data does not significantly support the key claim that extinction learning towards the CS+ occurred, but I do not feel it is acceptable to draw from the other measure for this effect alone. If the PSR measure is of primary/significant importance to support the validity of your paradigm, please consider adding all of these results to the main manuscript.*

We agree with this point and have moved the PSR analysis to the main manuscript. In addition, the SCR Results section no longer includes the PSR analyses, and clearly states the absence of a significant Stimulus x Time interaction effect in extinction ($p = 0.053$). For completeness, we additionally report trend-level post hoc tests showing CS+/CS- differentiation during early extinction but not during late extinction, consistent with an initial differentiation that attenuates across extinction training.

> *Subjective and (some) skin conductance responses do not completely support the success of the learning paradigm. For example, CS+/CS- differentiation in both subjective domains and SCRs persisted after extinction training. Can the authors comment on how this might influence the interpretation of their results more generally? What does it mean if these expectations do not appropriately translate to updated subjective appraisals in your participants, contrary to the model from which the parametric modulators were derived would predict?*

The persistence of CS+/CS- differentiation in self-reports after extinction, and the return of CS+/CS- differentiation in both self-reports and physiological measures during the recall test, is not unexpected. For self-reports administered after extinction, such persistent CS+/CS- differences are commonly observed in the human fear extinction literature (Hermans et al., 2006; see also Lipp et al., 2003), and may reflect that initial extinction learning establishes a new inhibitory association that suppresses, but does not erase, the original fear memory (Bouton, 2002). At recall on day 3, the remaining differentiation in both self-reports and physiological responses is consistent with spontaneous recovery, a well-documented phenomenon in extinction research (Bouton, 2002). As noted earlier (Reviewer 2, public review response: Extinction in self-reports and skin conductance responses (SCRs)), additional analyses showed that ratings were significantly reduced after extinction and recall compared to acquisition. Thus, while residual differentiation in self-reports remained after extinction and recall, its magnitude was diminished, indicating that extinction learning occurred but was incomplete. This pattern is consistent with partial updating of subjective appraisals in accordance with the reinforcement-learning model used to derive the parametric modulators, rather than a failure of updating.

> *Figures*
>
> *Figure 1: Please ensure that the summary of your results in the figure legend is consistent with the quantitative results reported. Example 1: "On day 2, there was a loss of differentiation during extinction training.", however, a significant effect of the stimulus, and time remained (but no interaction). Please tone down this interpretation, or make it clearer how the difference in the initial extinction trials was quantified. If the ANOVA-type analysis was only performed in the first half, this was not clear. Example 2: "During initial reacquisition, there were again differential responses to the CS+ and CS-, which decreased in reextinction and the unexpected US phase". I appreciate that you refer to the difference decreasing, rather than disappearing altogether, but the magnitude of this difference is not reported in the manuscript, and there does remain a significant difference in the amplitude.*

We thank the reviewer for this helpful feedback. We have revised the figure legends to tone down overly strong statements and ensure that all descriptions are in correspondence with the quantitative results. For clarity, we have also added significance markers for (trend-level) post hoc comparisons (CS+/CS- differentiation within early and late blocks for each phase) to revised Figures 2 and 3 displaying SCRs and PSRs.

> *Figure 2, 3, 4: I found it quite confusing to have uncorrected and corrected results displayed in the same way in the same figure. E.g. Figure 2A which, as far as I can tell shows trend-level results for the cerebellum, and corrected results for the VTA. For Figures 2 and 3 it was also not immediately clear which colour bar related to which map. Figure 4A appeared to be missing colour bars. I suggest the authors consider (as much as possible) standardising the colour bar scales, such that the maps across figures/sub-plots are more directly comparable, and differentiate more clearly between corrected and uncorrected results. The 3D renders in Figures 2 and 3 are a little hard to see - would it be possible to make it not so transparent?*

We use “trend-level” to refer to uncorrected voxelwise t-maps at $p < 0.05$, and “significant” to refer to TFCE/FWE-corrected effects at $p < 0.05$. This distinction was not sufficiently clear in the original figures. In the revised figures, uncorrected t-maps are displayed with a grey striped background frame. Colorbar scales were not standardized, as different panels display different statistical quantities (TFCE values versus t-values), and scaling was chosen to visualize variation within each contrast rather than enforce comparability across panels, which would have reduced interpretability. In addition, the missing colorbar in Figure 8A (formerly Figure 4A) has now been added; it matches the colorbar shown in Figure 8B. See also Reviewer 1, public review response, point (3) Clarity and Visualization of Results.

> *Is it possible to annotate significant effects on Figure 1 and Supplement Figure 1? The use of square markers makes it quite hard to tell the value of each point, which, given the small scale of the y-axis is quite important for interpretation. Could the authors consider remaking these plots with smaller dots?*

We have added post hoc significance markers to Figures 2 and 3 displaying SCRs and PSRs to facilitate interpretation. These markers reflect post hoc comparisons of CS+/CS- differentiation within early and late blocks. In cases where the Stimulus x Time interaction was not significant, the corresponding post hoc markers are still shown but are indicated in red to denote their trend-level status. In addition, the plots have been remade with smaller dots to make individual values clearer.

> *Discussion*
>
> *The authors state "Because aversive stimulus presentation results in pronounced cerebellar activations, we were unable to separate cerebellar activation related to the*

> *unexpected (initial acquisition trials) and the expected (late acquisition trials) presentation of the US." Could the authors compare between early[CS+>CS-] and late[CS+>CS-] acquisition (which I believe were created in the event-based analysis but results not reported), or between the first 3[CS+ with US>CS-] and later [CS+ with US>CS-] to assess this?*

In our terminology, the suggested comparisons (early vs. late [CS+ > CS-] or first three vs. last three [CS+ > CS-]) reflect changes in US prediction rather than prediction error. The statement in the Discussion refers specifically to cerebellar activation during US presentation, where distinguishing between expected and unexpected presentations is complicated by the strong cerebellar activation elicited by the electrical US itself. Moreover, when comparing early "unexpected" US presentations with later "expected" ones, the relatively higher activity in early trials could reflect habituation of the US sensation (i.e., non-associative learning) rather than a prediction error, making interpretation difficult.

Because the current manuscript focuses on reward-like prediction errors, we did not report these US prediction or presentation contrasts in detail. In brief, the suggested comparisons of early versus late CS-related differentiation (CS+ > CS-), revealed only limited trend-level activity. In contrast, US-related responses during acquisition showed robust activations in the cerebellar cortex, DCN, and VTA across the acquisition phase. Comparisons between the first three US presentations and later US presentations showed broadly distributed and stronger responses during early acquisition than during later US presentations. This pattern seems to be more consistent with non-associative effects, such as sensory habituation to the electrical stimulation, rather than with prediction-error–related processing. We have therefore not included them in the manuscript, but would be open to providing them in the Supplementary Information if the editor or reviewers consider them essential.

> *General*
>
> *In your pre-registered analysis plan you state "we will explore the use of DCM in a larger network that encompasses known constituents of the fear extinction network, in addition to the cerebellum and VTA.". You have plenty of results to discuss in the current manuscript and adding this may complicate the narrative, but that being said, please either perform and include this analysis as you proposed or explicitly mention why this was not completed. You could also consider adding a whole-brain activation map for the key phases of the experiment. Please also double-check other pre-registered points, for example - the sample size justification is also different.*

We decided not to include whole-brain DCM analyses in this manuscript and not to report whole-brain activation results extensively, as the study was primarily hypothesis-driven with a focus on cerebellum-VTA interactions. While we recognize that whole-brain analyses are of interest and plan to explore them in future work, they were considered outside the scope of the current paper. This deviation from the preregistration is now explicitly noted in the revised manuscript.

Regarding the sample size justification, the preregistration contained an error: the parameters were reported incorrectly. The correct sample size justification was already provided in the original 2019 grant application and is correctly reported in the current manuscript. The underlying power analysis was the same, but with different alpha levels depending on whether the study involved healthy participants (where larger samples are feasible) or rare patient populations (where stricter alpha levels are not practical). We have clarified this point in the manuscript under deviations from the preregistration.

Additional changes made in manuscript by authors

To provide a complete overview, we also note changes made independently of specific reviewer comments:

Methods

In the computational modeling section, “reextinction” was mistakenly mentioned where “reacquisition phase” was intended (the initial phase of the volatile phase before experience replay). This has been corrected.

The term “trial sequence” is used in computational modeling, whereas counterbalancing in the fear conditioning methods used different terminology. We added a clarifying sentence in the modeling section to make this consistent.

References in the pupil size analysis section (Jentsch et al. 2020; Mathôt et al. 2017) were misplaced and have now been moved earlier in the sentence.

The citation for MRIcroGL software was updated to the current Nature Methods reference.

We added a reference to Doubliez et al. 2025 which used the same three-day paradigm in a behavioral study showing similar physiological responses.

Supplementary information

During revision, we noted that the SCR statistics had been computed on an earlier preprocessed dataset version, whereas the finalized corrected dataset was already used for plotting and for estimating prediction and prediction-error values in the reinforcement-learning model. We therefore recomputed the SCR statistics on the finalized dataset for the sake of consistency; this did not change any main effects, interactions, or conclusions, with the only difference being an exploratory late-acquisition CS+/CS- post hoc shifting from non-significant to $p < 0.05$ (interaction still non-significant). Updated statistics are reported in the Supplementary information.

Post hoc significant differences in Table S3 are now marked in bold, as the formatting was missing previously.

To align behavioral analyses more closely with the event-based fMRI approach, we additionally examined physiological responses using a first three versus last three trial division within each phase. These analyses yielded patterns consistent with those obtained using the original early/late block division and are reported in the Supplementary Information.

We added a new supplementary figure (Figure S4) showing the location of the cerebellar VOI on a SUIT flatmap and added a corresponding cross-reference in the Methods section (Volumes of interest (VOI) definition)

References

Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. Biological Psychiatry, 52(10), 976–986. *https://doi.org/10.1016/S0006-3223(02)01546-9*

Bouton, M. E. (2004). Context and Behavioral Processes in Extinction: Table 1. Learning & Memory, 11(5), 485–494. *https://doi.org/10.1101/lm.78804*

Constantinou, E., Purves, K. L., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M., Craske, M. G., & Eley, T. C. (2021). Measuring fear: Association among different measures of fear learning. Journal of Behavior Therapy and Experimental Psychiatry, 70(September 2020), 101618. *https://doi.org/10.1016/j.jbtep.2020.101618*

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

Doubliez, A., Köster, K., Müntefering, L., Nio, E., Diekmann, N., Thieme, A., Albayrak, B., Nicksirat, S. A., Erdlenbruch, F., Batsikadze, G., Ernst, T. M., Cheng, S., Merz, C. J., & Timmann, D. (2025). Dopaminergic drugs modulate fear extinction-related processes in humans, but effects are mild. Brain Communications, 7(5), fcaf333. *https://doi.org/10.1093/braincomms/fcaf333*

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. *https://doi.org/10.7554/eLife.46831*

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. Biological Psychiatry, 60(4), 361–368. *https://doi.org/10.1016/j.biopsych.2005.10.006*

Kalisch, R., Gerlicher, A. M. V., & Duvarci, S. (2019). A Dopaminergic Basis for Fear Extinction. Trends in Cognitive Sciences, 23(4), 274–277. *https://doi.org/10.1016/j.tics.2019.01.013*

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? Learning and Motivation, 34(3), 219–239. *https://doi.org/10.1016/S0023-9690(03)00011-0*

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., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … 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. Neuroscience and Biobehavioral Reviews, 77, 247–285. *https://doi.org/10.1016/j.neubiorev.2017.02.026*

Milad, M. R., & Quirk, G. J. (2012). Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. Annual Review of Psychology, 63(1), 129–151. *https://doi.org/10.1146/annurev.psych.121208.131631*

Salinas-Hernández, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., & Duvarci, S. (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. ELife, 7, e38818. *https://doi.org/10.7554/eLife.38818*

Salinas-Hernández, X. I., Zafiri, D., Sigurdsson, T., & Duvarci, S. (2023). Functional architecture of dopamine neurons driving fear extinction learning. Neuron, 111(23), 3854-3870.e5. *https://doi.org/10.1016/j.neuron.2023.08.025*

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. Annual Review of Clinical Psychology, 9(March 2013), 215–248. *https://doi.org/10.1146/annurev-clinpsy-050212-185542*

https://doi.org/10.7554/eLife.105399.2.sa0