Resolving temporal threat uncertainty by observational learning involves the amygdala, hippocampus and anterior insula

Temporal uncertainty in observational learning

Simon Knobloch[a]*, Theresa, Jansen[a] , Lara Hille[a], Madeleine Müller[a,b], Leonie Rumpf[a]., Jan Haaker[a]

[a] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.

[b] Department of General Psychology and Cognitive Neuroscience, Charlotte Fresenius University, 20148 Hamburg, Germany

*Corresponding author. Email address: simon.knobloch@rwth-aachen.de

This manuscript is currently under review at *Translational Psychiatry*

Conflict of interest statement: The authors declare no competing financial interests.

Abstract

By relying on the observation of others' experiences, humans learn about threat while avoiding harmful experiences. Yet, previous neuroscience research has focused on observational threats that are predictable. While the neurobiological distinction between temporally predictable (cued) and unpredictable (contextual) threats has been well-characterized in firsthand learning. In this study, we developed a novel observational paradigm in which participants learned from predictable (P) and unpredictable (U) observational threats, as well as a no-threat (N) condition and encountered the same conditions during an expression phase based on the NPU paradigm to investigate how the brain encodes predictable and unpredictable threat cues observed in others. Participants in Experiment 1 (n=20, male and female) and Experiment 2 (n=23, male and female) successfully learned threat contingencies, showing heightened threat expectations for predictable cues and unpredictable contexts. This converged with neural (fMRI, Experiment 2) responses in the anterior insula during the expression phase. Reflecting the dynamic process of learning, the amygdala responded to predictable threat cues with a linear decrease across trials. Interestingly, we found that responses to others' pain was enhanced within the amygdala, insula and hippocampus, when participant could learn to predict threats, as compared to unpredictable conditions. Our findings suggest that humans learn to resolve temporal uncertainty, relying solely on observation, which thereby lays a foundation to the concept of fear and anxiety in social groups.

1. Introduction

Accurate prediction of threats is key to survival. Since we live embedded in sociocultural surroundings, observational threat learning enables us to predict dangers without any direct encounter of threats [1-3]. In observational fear learning it has been shown that the demonstrator's expression of pain serves as an unconditioned aversive stimulus that enables to learn a predictive threat cue, which was conditional for the response observed in the demonstrator [4]. The learning processes closely mirror firsthand threat learning both in behavioral fear reaction and neural response [4-7]. Specifically, the amygdala has been shown to play a central role in direct learning [8-11] and observational learning [4,5]. As a crucial difference between observational and direct learning, it has been noted, that in social learning the anterior insula, rather than the amygdala plays the central role in relaying the input of the observed unconditioned stimulus to a neural network for aversive learning [7].

Yet, in previous protocols of observational threat learning, the observed US was predictable by non-social signals (e.g., geometric shapes predicting the observed US).

From a functional perspective, humans have to be able to resolve information that indicate both predictable threats and threatening contexts or situations, in which threats occur unpredictably [4,12]. To date, it remains unclear, how unpredictable threats are learned by observation, which includes learning that is only dependent on social information. In fact, predictability of outcomes modulates the perception of aversive experiences not only for firsthand pain, but also when inferring the unpleasantness for others [13].

Regarding the distinction of predictable threat cues and context-related sustained threat, the NPU-paradigm (no-pain, predictable pain and unpredictable pain) has been developed to model responses comparable to the dissociable defense mechanisms of fear (towards predictable threats) and anxiety (towards contextually sustained threats) [14]. Participants learn to discriminate the three conditions and develop dissociable reactions to the three conditions, with an elevated anxious arousal (e.g., measured by startle-responses) towards the predictable cue and the unpredictable context [14,15]. Neurally, a differential contribution of the amygdala to predictable threat learning [8,15-17] and the hippocampus [8,18] to uncertain threat learning has been proposed to be a central discerning feature of the two threat types. It has to be noted though that the amygdala has also been shown to be involved in contextual (i.e. unpredictable) threat conditioning [18,19]. The insular cortex, apart from the mentioned role in observational learning, has been shown to be involved in the anticipation of uncertain threats [15,17,20-23].

In our study, we aimed to examine the acquisition of predictable threat cues and sustained responses to unpredictable threats by social means. To this end, we developed a novel paradigm, in which participants observe a demonstrator encountering threat (delivered through painful stimulation) in an NPU-based design. The cued pain occurrence can be learned to be resolved in the predictable condition but is temporally unresolvable in the unpredictable condition. We hypothesized (in Experiment 1) that participants will be able to resolve the threat-related information by observation and show discriminable threat expectations to predictable and unpredictable cues and contexts when confronted directly in the expression phase. We pre-registered our hypotheses in Experiment 2 that neural responses in the amygdala, hippocampus and insular cortex underlie learning from others' pain to resolve temporal predictability.

2. Methods

2. NPU paradigm

Conceptual basis of the experimental paradigm developed for this study was the NPU paradigm that includes firsthand experiences of an unpleasant electrotactile stimulation to induce a threat [14]. Our study adapted this protocol to examine observational learning within a predictable pain condition [P], unpredictable pain condition [U] and no pain condition [N]. Conceptual basis for our observational learning paradigm is a well-known two-phase experimental paradigm, in which participants learn threat/cue contingencies by observation in stage one (observation phase), and in stage two (expression phase) directly encounter the cues learned to predict threat or safety [12]. The paradigm enabled us to distinguish between the observation of a no-pain condition (N) and two conditions in which the outcome certainty was equal but varied with regards to the temporal resolvability of the pain for the demonstrator. Only in P, participants were able to learn when the pain would occur during the individual trials.

We report two experiments in this publication. In Experiment 1 we developed and tested the observational learning paradigm in the behavioral lab. In Experiment 2 we adapted the paradigm for use in fMRI. We, thus, were able to retest preregistered behavioral results from Experiment 1 in an independent sample in Experiment 2 and advance the understanding of the underlying mechanisms through the analysis of neural correlates. The conducted analyses were preregistered prior to data-analysis [24].

2.2 Experiment 1

2.2.1 Participants and procedure

20 healthy participants (see Table 1 for descriptives) were recruited by online and black-board postings at the University Medical Center Hamburg. All participants were fluent in German and able to understand the task instructions. Further inclusion criteria were an age between 18 and 40, sufficient general health and normal or corrected to normal vision. Exclusion criteria were current or previous psychiatric disorder, current use of psychoactive medication or substances and current or previous neurological disease. Participants that had already participated in threat-conditioning studies at our institute within three months preceding the participation were also excluded. Eligibility criteria were assessed in structured interviews via telephone and verified on site by a physician. Study data was collected in a single

appointment of approximately 90 minutes. After giving informed consent to participate in the study, participants filled out a questionnaire collecting sociodemographic data.
The experiment consisted of two blocks of 25 minutes (see 2.2.3 Experimental Design) and was conducted on a computer in the behavioral lab. After completing the experimental paradigm participants filled in a post-experimental questionnaire, in which they answered questions about the cue/pain contingencies and rated the observed pain expression with regard to naturality and their empathy towards the demonstrator. All questionnaires were conducted via soSci (soscisurvey.de, München, Germany).
Study procedures were in line with the Declaration of Helsinki. The ethics committee of the Hamburg medical chamber approved the study. Participants that finished the study received an expense compensation of 15€.

2.2.2 Experimental design

The experiment was divided into two phases: observation-phase and expression-phase.
In the observation-phase participants were instructed to watch videoclips of a novel blood-pressure measurement approach, which was highly precise yet painful. They were told to watch the patient's pain-expression as well as the screen, visible in the video. Participants were not instructed about the condition-design and cue/pain contingencies. The screen in the video showed a matrix of 3 horizontal grey bars (rectangles) stacked vertically. Each bar had a colored portion of different size and position and a black slider at the left-hand side of the bar. At the start of a trial, one of three sliders started moving from the left-hand side of the bar to the right. In the P condition, the pain expression of the demonstrator was predictable by the slider crossing the colored part of the bar. In U, the slider position had no predictive value for the pain expressed by the demonstrator, i.e. the demonstrator's pain responses were visible at random slider positions in each video clip. In the N condition, no pain expression was visible (see figure 1 b for an overview of slider/pain relation), while the slider was crossing the entire bar. Each video-clip lasted from the moment the slider started moving till it reaching its final position on the right-hand side of the grey bar (24s). Participants were asked to rate the pain intensity for each clip. In total participants completed 12 trials in each condition.

In the second phase, the expression-phase, participants were directly confronted with what was previously visible on the demonstrator's screen (as depicted in Fig 1 b). Similarly, the trials lasted from the moment the slider started moving till it reached its final position on the right-hand side of the grey bar (24s). No pain was administered to the participants.

Participants were asked to answer questions regarding their expectations about painful stimulation in each clip at the condition onset (i.e., context onset) and when the slider crossed the colored portion (i.e., cue). In total participants completed 12 trials in each condition. The inter-trial interval included a fixation cross that was displayed for 5 seconds.

Ratings

During the observation phase, participants were asked to rate the observed pain intensity on a scale generating values from 1-100 (1 = no pain, 100 = strong pain). The rating query appeared during the pain-expression of the demonstrator. For the rating, the video-clip was paused and blurred. In the N condition, the participants rated the same question while no pain was visible.

During the expression phase, participants were asked to rate their threat expectation and certainty at three locations of the moving slider (context-onset-expectation: slider position 2,5 seconds before the colored part of the bar, cue-expectation: slider position in the middle of the colored part of the bar, context-offset-expectation: slider position 2.5 seconds after the colored part of the bar). Note that due to a technical error in the script the first four participants were not asked for context-offset-expectation (context-offset-expectation n = 16). Answers were given on a scale generating values from -100 (certainly no pain) via 0 (uncertain) to 100 (certainly pain). For each expectation-query, participants were asked to rate both an expectation for the demonstrator confronted with that screen setup and a self-rating when picturing themselves being confronted with the screen setup.
All answers were generated by using the arrow-keys of a standard keyboard.

2.2.3 Stimuli

33 Videoclips for P and U with randomized pain onset and -duration were recorded. The videos show a demonstrator model who is watching a screen and receives painful heat-stimulation. The demonstrator wears a cuff around his right lower forearm (see Fig 1 a) and the heat stimulation device (via a thermode) is placed under the cuff. The response of the demonstrator to the heat stimulation reflected the painful experience. In order to choose video clips for the experiment that were comparable in pain expression, two independent individuals rated five components of the demonstrator’s pain expression: lip press, head protrusion, eye closure, forehead wrinkle and masseter contraction. Additionally, three individuals rated

independently the binary presence of the five components as well as pain visibility duration (in seconds) and pain-intensity (scale from 1-10, i.e. no pain resp. very strong pain). All clips that featured at least four pain components (marked as present by at least one of the three reviewers) entered a propensity-score based matching. Propensity score was calculated by the mean rated pain-intensity and duration. The caliper width of the matching was reduced until a subset of 12 clips for each condition remained for Experiment 1. The 12 included clips (Table 2 describes mean and range for pain-, cue duration and pain/cue overlap) did not differ significantly in pain-intensity (p =.690) and visible pain duration (p = .918) in Experiment 1.

2.3 Experiment 2

2.3.1 Participants and procedure

GPower 3.1.9.7 (RRID:SCR_013726) was used to estimate the sample size for Experiment 2. Based on the pilot study (N=20), the relevant interaction yielded an effect size of partial eta squared = 0.331, which results in a power of >99% for 20 participants, given an alpha rate of 0.05. Thus, we determined the sample here to be 25 participants including dropouts. Thus, 25 healthy participants (see Table 3 for descriptives) were recruited by online and black-board postings at the University Medical Center Hamburg. Eligibility criteria were equal to Experiment 1. Additionally, participants unable to undergo fMRI scanning due to ferromagnetic or electrical implants as well as unspecified foreign bodies were excluded from the study-participation.

Study data was collected in a single appointment of approximately 120 minutes. The procedure was similar to Experiment 1 except for Experiment 2 being conducted in the fMRI scanner. In Experiment 2 skin conductance response [SCR] was collected to assess the psychophysiological response. Furthermore, in Experiment 2 participants filled out the State-Trait Anxiety Index [STAI] [25] and the Affective and Cognitive Measure of Empathy [ACME] [26].

Neuroimaging acquisition consisted of two functional imaging blocks, each taking approximately 25 minutes (observation and expression, see 2.3.3 Experimental Design). Subsequently a T1 weighted image (ca. 5 min.) for co-registration and a T2 weighted image (ca. 10 min.) were collected. The data from the T2 weighted image is not reported in this study.

2.3.2 Experimental design

The two-phase experimental design did not differ substantially from Experiment 1. In Experiment 2 participants rated pain intensity in the observation phase and threat expectations only in three of the twelve trials. The other nine trials ran without rating interruption for the analysis of neural correlates.

In contrast to Experiment 1, phase participants wore an identical cuff , as seen on the demonstrator, in Experiment 2 and were told that it might inflict some unpleasant yet not highly painful stimulation during the expression phase. As in the observation-phase, participants rated only one identical trial three times. The remaining 9 trials ran without rating interruption. Furthermore, participants completed one rating trial of the expression-phase before the observation phase to collect baseline expectations without relation to the information learned in the observation-phase.

All answers were collected by using a 4-button box. The scales used in Experiment 2 were identical to the ones from Experiment 1. In the expression-phase, participants were only asked to rate threat expectation and certainty regarding themselves being in the same situation as the demonstrator before. No expectations about the demonstrator were collected.

2.3.3 Stimuli

For Experiment 2, caliper width was further reduced to reduce the subset to 10 video clips. The 10 included clips did not differ significantly in pain-intensity ($p= .612$) and visible pain duration ($p= .912$) in both subsets respectively (Table 2 describes mean and range for pain-, cue duration and pain/cue overlap). One clip of P and U respectively, matched in pain-intensity and -duration was chosen and used thrice for the three rating trials. The other 9 trials were used once each for the non-rating trials.

2.3.5 Image acquisition

All images were acquired using a 3 Tesla Siemens Prisma scanner (Siemens AG; München, Germany) at the Institute for Systems Neuroscience at the University Medical Center Hamburg-Eppendorf. For this study, echo-planar imaging [EPI] sequences sensitive to blood oxygenation level-dependent changes were acquired for the functional analysis. 52 axial slices

of 1.5mmx1.5mmx1.5mm voxels were acquired (TR = 1848ms, TE = 30ms, flipangle = 70°). Subsequently, an anatomical image was acquired (T1 weighted 3d image, magnetization prepared rapid acquisition gradient echo image, 1mm x 1mm x 1mm voxel size, TR = 7.1ms, TE = 2.98ms, TI = 1100ms, matrix = 256 × 256, 240 slices, flip angle = 9°).

2.3.6 SCR

SCR was collected using two palmar electrodes on the participants left hand. Spike2 software (Cambridge Electronic Design Limited, Cambridge, England, RRID:SCR_000903) was used for data collection. MRI-pulses were collected to synchronize the experiment's logfiles.

2.4 Statistical analysis

2.4.1 Behavioral analysis

In both Experiments (1&2) mean threat expectation ratings before and during the cue from the expression phase were analyzed in a 3x2 repeated measurement ANOVA with condition (P, U, N) and cue/context-onset (cue-expectation, context-onset-expectation) as factors. For post-hoc comparison we aimed to characterize an interaction between condition and cue vs. context. To that end, the 'cue-prominence' was calculated for all conditions separately by subtraction of the context-onset-expectation from the cue-expectation. Cue prominence was compared among the conditions in paired-sample t-tests (P vs. N, U vs. N, P vs. U). In Experiment 1 a similar repeated measurement ANOVA and post-hoc t-tests were calculated for expectation of pain for the demonstrator. To account for beliefs held by participants unrelated to the observable predictability of the conditions (i.e., baseline), we report pre-observation ratings for threat expectation and compare them similarly across conditions in Experiment 2. Additionally, context-offset-expectation (slider-position behind colored part of the bar) were compared in a 1x3 repeated measurement ANOVA among the conditions in both Experiment 1 and 2. Significant effects were followed up using paired-sample t-tests. Furthermore, mean pain ratings from the observation phase were compared among the pain-conditions in a paired-sample t-test (P vs. U).

Alpha-level was set to $p < .05$. Multiple comparisons were accounted for through Bonferroni-Holm correction. All statistical analyses were conducted using JASP 0.18.3.0 (RRID:SCR_015823). Matlab 2023b (Mathworks Inc., Natick, Massachusetts, USA, RRID:SCR_001622) was used to read-out logfiles and count the participants' answers.

2.4.2 fMRI analysis

Participants were checked for excessive movement prior to analysis (preregistered exclusion criterium = 3x voxel-size lead to deletion of affected volumes, participants exceeding 25% data quality issues were excluded from analysis). For preprocessing EPIs were realigned to correct for motion and the six resulting movement parameters were included in the first-level analysis. Subsequently, images were coregistered to the anatomical image (reference: T1, source: EPI) and normalized into MNI space using SPM 25 defaults.

In both experimental phases of Experiment 2 (observation and expression) cue and context regressors were modelled following the focus of the behavioral analysis of cue- and context-onset-expectation. To prevent cue-, context- and pain-regressors from being colinear, context-onset-regressors were modelled as event regressors at the onset of slider movement and cue regressors as event regressors in the moment the slider started the passage of the colored part of the bar. Pain-regressors were modelled as event-regressors at the onset of pain-expression, as rated by three lab members for each video clip (mean of rated onsets). For N, whose stimulus videos did not contain visible pain in the demonstrator, a no-pain event regressor was modeled starting approximately (in a random interval) 2s after the cue-regressor to allow to compare social information processing (pain>no pain contrasts). A box-car regressor was modelled for the whole 24 s of each videoclip representing sustained context activations through the experimental condition. ITI, instructions and ratings were modeled as separate box-car regressors. Regressors were convolved with the hemodynamic response function in the first level analysis. Temporal autocorrelations were accounted for through AR1. The resulting beta images were smoothed using a 6mm FWHM Gaussian kernel.

For the second-level analysis a one-factor (3 levels: experimental condition) full factorial design was calculated for each regressor separately: cue, context-onset and context, both in observation and the expression phase. Furthermore, a two-factor full-factorial design with one factor comprising cue and context-onset and the other factor the three experimental conditions was applied to analyze neural substrates of cue-prominence. We report the following, pre-registered, t-contrasts for the one-factor analyses: P>N, U>N, P>U, U>P. For the cue-prominence analyses, we report a pre-registered interaction contrast with cue >context-onset in P and cue < context-onset in U. For the analysis of pain related activation, we report a secondary t-contrast P&U>N for the analysis of neural correlates of pain-expression in

general. In a subsequent analysis (following Marschner and colleagues' analysis [8], the same first-level regressors were used with linear time-modification over the course of the trials, as pre-registered. We report the contrast P>U (i.e. increasing activation in P and decreasing activation in U) and U>P (i.e. vv.).

For all analyses we report activation of 3 predefined ROIs (right and left separately): amygdala, hippocampus, insular cortex from the Harvard - Oxford Cortical Structural Atlas (RRID:SCR_001476). Clusters that surpass peak-voxel FWE correction at $p<.05$ in small volume correction (across the three ROIs bilaterally) are reported. For amygdala clusters, ROIs of the basolateral and centromedial subregions of the amygdala [BLA and CMA respectively] were used to activations in mechanistically discrete amygdala subregions. Single-sided BLA and CMA masks [27] were generated using the Anatomy toolbox (V3, RRID:SCR_013273). SPM25 (RRID:SCR_007037) was used for the preprocessing and analysis of the fMRI data.

2.4.3 SCR analysis

The collected SCR data from the experiment was analyzed using PSPM7.0. Raw data was trimmed from the onset of the first non-rating trials till the last MRI-pulse of the two experimental phases separately. Trimmed data was assessed visually for data quality. Participants with at least 50% noisy- or zero-line data were excluded from further analysis. Regressors for context onsets, cue onsets (and pain onsets in the observation phase) were modeled and convolved with the canonical skin conductance response function. Following the design of the behavioral analysis, a 2x3 repeated measurement ANOVA with condition (3 levels) and cue/context (2 levels) as factors are reported. Significant factors were followed up using paired-sample t-tests. Pain related skin-conductance alterations were compared using paired-sample t-tests among the three conditions.

3. Results

3.1 Sample characteristics

In Experiment 1, the 20 recruited participants all entered the analysis. A description of the study sample can be found in Table 1.

In Experiment 2, 25 participants were recruited. One was excluded due to technical problems with stimulus presentation and erroneous data collection. A second participant was excluded,

due to excessive movement during acquisition preventing the generation of a usable T1 image for coregistration.

3.2. Behavioral results

3.2.2 Experiment 1

First, we tested the central hypothesis that participants learn the temporal predictability of cues the aversive outcomes by observation. To this end, we tested participants' threat expectation for firsthand outcomes in P vs. U in the expression phase. Indeed, the repeated-measurement ANOVA of rated threat expectation revealed a significant influence of the condition ($F = 15.175$, $df = 2$, $p < .001$), cue/context-onset ($F = 62.033$, $df = 1$, $p < .001$) and their interaction ($F = 9.383$, $df =2$, $p < .001$) (see Table 4). Post-hoc t-tests to follow upon this interaction compared cue and context expectation, i.e., 'cue-prominence' (cue-expectation – context-onset-expectation) between conditions thus revealed higher cue-related expectation in P ($t= 3.994$, $df = 19$, $p = 0.00078$, $p_{Holm} = 0.00234$) and U ($t= 2.449$, $df = 19$, $p = 0.024$, $p_{Holm} = 0.048$) as compared to N. The comparison of the cue-prominence between P and U revealed a trend of a higher cue-prominence in P ($t= 1.997$, $df = 19$, $p = 0.06$, $p_{Holm} = 0.06$). Table 5 and Table 6 give an overview of the behavioral outcome measures and the statistical analyses. Plots of the behavioral results can be found in Figure 2. We conducted an additional analysis comparing cue and mean context expectation (mean(context-onset, context-offset)) using the same analytic approach, revealing similar results of higher cue prominence in P and U, compared to N, as well as higher in P compared to U (see supplementary table S1-S3).

We additionally asked for participants expectation about the pain that the demonstrator might receive during the expression phase, which did not match the learned threat contingencies from the observation phase. You can find this additional analysis in the supplements (supplementary text S1).

In order to test if threat expectation were sustained, we compared threat expectation after the cursor passed the colored cue, across the conditions, which however only revealed a trend-wise effect ($F = 2.772$, $df = 2$, $p =0.079$, 1x3 ANOVA). Nevertheless, participants descriptively rated the expected pattern, with lowest threat expectation for N (mean ± SD: -0.82 ± 0.31), a medium threat expectation for P (mean ± SD: -0.66 ± 0.30) and the highest, sustained expectation of an aversive outcome in U (mean ± SD: -0.53 ± 0.43).

Participants furthermore rated trend-wise higher pain-intensity for the demonstrator in P as compared to U ($t= 1.910$ $df = 19$, $p = 0.071$).

3.2.2 Experiment 2

In Experiment 2 we were able to reproduce and advance the results from Experiment 1. Similar to Experiment 1, the repeated measurement ANOVA of threat expectation in the expression phase revealed a significant main effect of condition ($F = 22.642$, $df = 2$, $p < .001$), cue/context-onset ($F = 32.291$, $df = 1$, $p < .001$) and their interaction ($F = 13.158$, $df = 2$, $p < .001$) (see Table 7). Pre-registered post-hoc comparisons again revealed that the cue-prominence was increased in P ($t = 4.443$, $df = 22$, $p = 0.0002$, $p_{Holm} = 0.0006$) and U ($t = 2.814$, $df = 22$, $p = 0.01$, $p_{Holm} = 0.02$) in comparison to N and we found an increased cue-prominence in P as compared to U ($t = 2.808$, $df = 22$, $p = 0.01$, $p_{Holm} = 0.01$) (see Table 8, see figure 2). We found comparable threat expectation for the cue in P and U, but an increased threat expectation at context-onset in U as compared to P (Cue-Expectation P vs. U: $t = 0.891$, $df = 22$, $p = .382$, $p_{Holm} = .382$; Context-onset-Expectation = U vs. P: $t = 2.615$, $df = 22$, $p = .016$, $p_{Holm} = .032$). As expected, baseline expectation before observational threat learning were not different (see Table 9). Similar to analyses of Experiment 2, we conducted an additional analysis comparing cue and mean context expectation (mean(context-onset, context-offset), see supplementary tables S4-S6).

Participants furthermore sustained a differential threat expectation across conditions after the cursor passed the colored cue ($F = 5.989$, $df = 2$, $p = .005$). Similar to the pattern found in Experiment 1, we found an increased rating for both P ($t = 3.941$, $df = 22$, $p = .00066$, $p_{Holm} = .002$) and U ($t = 2.581$, $df = 22$, $p = 0.017$, $p_{Holm} = .034$) as compared to N, however, no difference was revealed between P and U ($t = 0.313$, $df = 22$, $p = .758$, $p_{Holm} = .758$), which might indicate that the predictable condition was evaluated as potentially dangerous, even when the predictable danger signal passed.

In contrast to our pre-registered hypothesis, we found no difference in ratings for the demonstrator's inferred pain between P and U ($t = 1.384$, $df = 22$, $p = .180$). Table 10 gives an overview of the behavioral measures taken in Experiment 2 descriptively.

3.3 fMRI results

3.3.1 Observation phase

Resolving cues and contexts during observational threat learning

In line with meta-analytic evidence [28,29] we found that the observation of pain across P and U elicited activation in the left insula ($t = 5.93$, $df = [1,66]$, MNI = [-36, 18.5, 0.5], p(FWE) = 0.00006) and amygdala ($t = 4.54$, $df = [1,66]$, MNI = [-28.5, 0.5, -22], p(FWE) = 0.0031; see Figure 3 a) – f)), when compared to no observed pain. Interestingly, observation of

predictable pain (P>U) was further encompassed by higher parameter estimates in the right hippocampus (t = 4.27, df = [1,66], MNI = [33, -32.5, -10], p(FWE) = 0.0145), when compared to observed pain that was unpredictable (see Figure 3 g) – i)).

Next, we tested for neural activation that would later reflects the participant`s ability to distinguish between cue and context-onset in different conditions of temporal predictability. Specifically, we contrasted (pre-registered) hemodynamic responses reflecting cue prominence (P: Cue > Context; U: Context > Cue) between P and U, that mirror participants rated higher cue prominence for P, as compared U in the expression phase. This contrast only revealed a cluster trending towards FWE-corrected significance in the left (anterior) hippocampus (t = 3.66, df = [1,132], MNI = [-21, -20.5, -20.5], p(FWE) = 0.056) (see Figure 4 a) – c)).

In order to account for the learning-dependent changes during observational learning, we modelled linearly changing activity across trials, as preregistered and in previous examinations of firsthand cue and context learning [8]. The interaction-contrasts reflecting the rated cue-prominence (contrast: linear decrease cue: P > U; linear decrease context U> P) revealed activation in the the right insula (t = 3.76, df = [1,132], MNI = [45, 6.5, -7], p(FWE) = 0.045), which reflected decreasing responses across learning to valid threat predictors, i.e. predictable cues and unpredictable contexts see [8]. As such, activation to predictable cues and unpredictable context-onsets that decreased across learning, while activity to unpredictable cues and predictable context-onsets was sustained or increased (see Figure 7 a) – c)). The similar pattern was found in the amygdala, specifically the bilateral CMA (t = 4.21, df = [1,66], MNI = [-18, -7, -13], p(FWE) = 0.009; t = 3.97, df = [1,66], MNI = [21, -5.5, -13], p(FWE) = 0.018) (see Figure 8 a) – c)). Here, activation decreased linearly across trials during observational learning of predictable cues as compared to sustained activity to unpredictable threat cues (contrast: linear decrease P > U), which might reflect an early integration of socially derived threat information in the predictable condition.

Cue- specific responses during observational threat learning

To examine only cue-related activation, we contrasted activation between predictable cues of observed pain, as compared to neutral cues (P > N), which revealed activation in the posterior part of the left hippocampus (t = 3.90, df = [1,66], MNI = [-13.5, -38.5, 2], p(FWE) = 0.041; see Figure 5). No differences in activation in our ROIs emerged in the comparison of

predictable and unpredictable cues (P > U), similar to the comparable reports of threat expectation.

Context- specific responses during observational threat learning

In the unpredictable condition, we found context-onset related activation in the left anterior insula (t = 4.51, df = [1,66], MNI = [-36, 11, -14.5], p(FWE) = 0.007) as compared to the neutral condition (U > N) (see Figure 6 a) -c)). We found no differences in activation in the ROIs at context-onset between U and P. However, the same contrast revealed sustained activation across the whole observational trial in right BLA (t = 3.77, df = [1,66], MNI = [21, -4, -20.5], p(FWE) = 0.027; t = 3.69, df = [1,66], MNI = [30, -4, -16], p(FWE) = 0.034) during the unpredictable threat condition, as compared to the predictable condition (U > P) (see Figure 3 d) -f)).

3.3.2 Expression phase

Resolving cues and contexts during observational threat learning

The neural activation during the expression phase was thought to reflect participant's ratings of socially shaped threat expectation that is higher for the cue (as compared to the context) in the predictable, as compared to the unpredictable condition (i.e., cue-prominence). This preregistered interaction contrast (P: cue > context-onset, U: context-onset > cue) revealed a cluster in the right insula (t = 4.23, df = [1,132], MNI = [34.5, 20, 2], p(FWE) = 0.009) (see Figure 4 d) – f)) and thereby suggests a neuronal activation pattern that reflects the participant's threat expectation to resolve temporal threat uncertainty.

To account for changes across trials in the direct expression phase (i.e., no firsthand outcomes), we modelled linearly changing activity, as preregistered. This analysis revealed activity in amygdala subregions, the CMA and BLA that diverged between predictable cues and unpredictable contexts (t = 4.56, df = [1,132], MNI = [24, -8.5, -14.5], p(FWE) = 0.00013), as specified in the cue-prominence interaction contrast (contrast: linear decrease cue: P > U; linear decrease context U> P). Again, activation decreased to predictable cues (as compared to predictable contexts) and unpredictable contexts (as compared to unpredictable cues) (see Figure 7 d) – f)).

Cue- specific responses during observational threat learning

In line with the linearly changing responses as found in the cue-prominence, we found transient CMA (t = 3.75, df = [1,66], MNI = [25.5, -8.5, -13], p(FWE) = 0.029) activation to

predictable cues in the expression phase (see Figure 8 d) – f)) as compared to increasing responses to unpredictable threat cues. This pattern is probably also driven by the retrieval of socially learned expectation when firsthand outcomes are lacking.

Against our preregistered hypothesis, the mean (not changing) responses to predictable threat-cues lead to no activation in the predefined ROIs. The comparison between unpredictable cues to predictable cues (U > P) revealed activation in the left hippocampus (t = 5.39, df = [1,66], MNI = [-19.5, -10, -23.5], p(FWE) = 0.0004).

Context- specific responses during observational threat learning

In line with finding on hippocampal-dependent contextual threat learning, we found that the hippocampus (t = 3.94, df = [1,66], MNI = [-24, -10, -28], p(FWE) = 0.032) in the expression phase engaged in sustained contextual activation across the whole trial (block regressor) in the unpredictable condition as compared to the predictable condition (contrast: linear decrease U > P) (see Figure 8 g) – i)). We neither found mean (not changing) activation in the ROIs at context-onset nor sustained contextual activation.

3.4. SCR-Analysis

Analysis of SCR data was prevented by bad data quality during data acquisition. Only 13 participants had valid SCR data for both experimental phases in Experiment 2, severely limiting inferentially validity. See supplement (supplementary text S2) for the analysis.

4. Discussion

Our results show that participants learned to resolve temporal threat uncertainties by observing painful experience in others to predicable cues and unpredictable context-onset. Expression of socially learned threat expectation that differ between predictable cues and unpredictable context are reflected by neural activation in the anterior insula and amygdala, that were already dynamically engaged during observational threat learning.

Mirroring the temporally dynamics of learning, we generally found decreasing activation to signals that were learned as valid predictors of threats: i.e., bilateral amygdala activity decreased to predictable cues during observational learning, whereas responses increased to unpredictable cues across trials. In uncertain learning situation sustained hippocampus activation was found across trials. Interestingly, we found that neural responses with key regions for observed pain, the amygdala and insula [28,29] were enhanced when the observation of pain could be used for learning to resolve temporal unpredictability (i.e., predictable threat cues), as compared to observed pain that was temporally unpredictable i.e., not learning relevant.

A differential contribution of the amygdala to cued threat learning, i.e. fear [8,16], and the hippocampus [8,15] to uncertain threat learning has been proposed to discern fear and anxiety neurobiologically. Furthermore, the role of the insular cortex in uncertain threat anticipation has been highlighted in studies that employed both, temporally predictable and unpredictable conditions [15,20,22,23]. Following studies have uncovered a less clear pattern of differentiation, though [19,21,23,30]. Our data does not fully support a clear mechanistic separation between brain regions. We found the right amygdala to be sustainedly activated during observation of unpredictable threats. The amygdala has been reported to react phasically to threat cues [15,16] or unpredictable context onsets [19,23] and threat confrontation [31], in paradigms with firsthand aversive experiences. However, sustained activation of the amygdala during unpredictable threat-contexts has been described to firsthand electric shocks [30]. Interestingly, connectivity analysis supports the view of a contribution of the amygdala towards the resolving of unpredictable threat contexts [32]. The sustained activation of the BLA, key region of associative threat learning [33,34] during unpredictable contexts might result from the ongoing yet futile attempts to integrate threat information. Against our expectation, our analyses revealed the hippocampus to be involved in both, encoding threat cues and uncertain threat contexts in observational threat learning, a seemingly contradictory finding to the proposed differential role in firsthand threat learning [8,15], but see [21,35]. In our study the hippocampus

reflected the direct comparison between cue and context in predictable vs. unpredictable conditions (i.e., the threat cue-prominence, as delivered subsequently rated in the expression phase) already during observation (FWE trendwise significant cluster). This role of the hippocampus in observational threat learning might be driven by an inherent aspect of observation, which is the uncertainty of firsthand outcomes for the participants. Hence, the hippocampal involvement during observational learning might reflect threat surveillance for own firsthand outcomes, rather than only temporal predictions for the demonstrator. We found that the threat information was resolved in the left anterior hippocampus, which is in line with the functional topography of the human hippocampus, where the L1 segment is associated with fear and conditioning and strongly connected to the limbic system [36]. The hippocampus's brief activation to predictable cues on the other hand took place in the left posterior portion, an area primarily associated with spatial perception but with strong functional links to the left IFG involved in semantic encoding, a plausible role during learning cued threat contingencies [36]. In line with previous research employing direct learning [17,20,22] the anterior insula was involved during observation of uncertain threat anticipation (i.e., unpredictable context onsets). Furthermore, the anterior insula was involved in resolving both contextual and cue-related threat information during the expression phase. These findings extend previous work employing firsthand threat learning , which fits the finding that the insular cortex has been highlighted as a highly connected hub in threat-learning [32] that computes personal affects in uncertain threat-situations [22]. In conclusion our data provide evidence that observational learning of predictable and uncertain threats is resolved in areas comparable to direct learning. In line with assumptions from direct learning [21], our data suggests that the neural systems involved in threat processing react to certain and uncertain threats with a difference in degree but not in kind. Previous research has highlighted the interplay of amygdala and hippocampus in the formation of complex emotional memories guiding our reactions to threat [37].

It has been suggested early on to accounts for learning dynamics during the acquisition of fear conditioning, i.e. dynamic rather than rigid neural responses to threat information [8,10,38]. Our study enables the comparison between a simple, predictable learning condition (conditioning (cued threat) and an unpredictable condition in which aversive outcomes are embedded in a rich but arbitrary environment (contextual threat). Consistent with previous findings [8], we observe a linear decrease in bilateral amygdala activation towards predictable cues, while responses to unpredictable cues increase over time. This pattern persisted in the expression

phase: predictable cues briefly activate the right amygdala, whereas unpredictable cues elicit sustained and increasing right amygdala activity. Similarly, the left hippocampus showed sustained activation in response to unpredictable cues. Interestingly, this explanation based on classical learning-theory, can contribute to the explanation of the interaction contrasts of cue prominence as well. During observation, the right anterior insula fades to activate to the easily resolved predictable cue and unpredictable context onset, whose meaning is more easily absorbed than v.v. For the expression phase a similar pattern can be observed in the right amygdala. It has been argued that the amygdala [8] as well as the anterior hippocampus [39] activate to novel information that has to be integrated. Our data fits this assumption well, with a limited amygdala response to predictable threat stimuli and a dually sustained (across length of individual trial & across trials) anterior hippocampal activation to unpredictable threats. We suggest that the predictable condition lead to rapid learning and quick resolution of prediction error, while unpredictability maintains prediction error throughout e.g. as modelled by: [40]. The linearly decrease in activation to predictable cues was primarily visible in the CMA, a region with strong structural links to brainstem regions to steer defensive behavior [34,41].The learning dynamics explained above might crucially also result from the translation of learned information into defensive behavior. Participants were able quickly to understand a predictable threat for others rather than themselves, quickly decreasing the necessity to link threat perception to defensive behavior.

We found that the left BLA and anterior insula were activated during the observation of pain in others, which is in line with previous studies showing both, faces in pain, and depicting noxious stimulations of others' extremities [29]. Interestingly, we found differences in neural responses to others' pain, based on the learning condition: We found higher hemodynamic responses in the anterior insular cortex and the BLAto observation of others' pain in the simpler learning condition (i.e. predictable condition), as compared to the unpredictable condition, where observed pain was not informative to learn threat signals [28,29]. In the unpredictable condition participants relied solely on social information, i.e. facial expression and gestures to infer pain and, thus, threat towards others. Our data nevertheless suggests that the prediction of social observation of pain by non-social cues enhanced neural responses, despite the seemingly (more) important social information in the unpredictable condition. Our results thereby imply the idea that the observation of predictable, as compared to unpredictable pain, enhanced neural responses (in the amygdala and insula) as a function of learning and that activation of the right hippocampus that might reflect the integration of the

observed pain into the learned concept. Previous research has described aberrations in vicarious stress processing to be linked to PTSD [42], highlighting the importance to better understand the neural mechanisms when observed pain is integrated and when it is ignored. A better understanding of mechanism influencing the perception of pain in others might also aid a better understanding of patient/provider interaction influencing e.g social placebo effects [43].

The original NPU-task used startle-probes to assess whether participants learned to discriminate between cue and context related threat contingencies [14] whereas we used SCR and explicit expectation queries. In this context, we define cue prominence as the difference of expectation or skin conductance measure with the cue present and at context onset. Previous research has successfully used SCR in direct NPU-paradigm [15,21] and observational learning [4,12]. Surprisingly we did not find meaningful psychophysiological reactions mirroring the explicitly given threat expectations. Potentially bad SCR data quality, which lead to a high drop-out rate in the analysis, has prevented the assessment of psychophysiological reactions in our study. Furthermore, a complex and inherently 'threatening' medical visual context in all conditions might have driven a strong condition independent SCR response. Future research should try to fill this gap, as previous research hints at robust SCR effects to observed pain and encounters of threat cues in the expression phase [4,12].

5. Limitations

We established a novel paradigm and both experiments conceptually resemble each other closely but slightly differed in the rating procedure. Furthermore, the threat expectations after cue presence, i.e. safety ratings, did not reflect discernable (by condition) effects in participants. In Experiment 1, the participants threat expectations for pain in others did not match the correctly transferred expectations of own threat. The query of expectations towards pain for the demonstrator and pain for the observer in quick sucession might have confused participants and was, thus, removed for experiment 2. Since observational pain decline was slow, this might have prevented the end of the colored bar to serve as a predictable safety cue. Our sample size for Experiment 2 might have been small, given the signal to noise limitations of fMRI research. Our implication of physiological signals in this study is limited, since we were unable to find psychophysiological responses to observed pain and to threat cues and contexts, which might be a result of generally bad SCR data-quality.

6. Conclusion

Here, we show the behavioral effects and neural basis of observational learning in temporally predictable and unpredictable threat-situations across two experiments. We were able to show that healthy participants are able to resolve the cue and context related threat contingencies by observing others' experiences and make inferences towards threat for themselves. Our results provide the neural basis for vicarious learning of cued and sustained threats, to build a basis to understand anxiety and fear, in humans as socially connected individuals. We were able to show that learning processes related to temporal certainty of threats closely mirror those from firsthand learning. On a neural level, our findings provide two novel perspectives, one that extends the neural mechanisms recruited to learn uncertain threat contingencies, and the second suggesting that such learning changes how we respond to vicariously observed pain of others. There is a solid evidence for firsthand threat learning mechanisms being linked to anxiety [44] and obsessive compulsive disorders [45]. A better understanding of how vicarious processes contribute to threat learning might aid broadening the understanding of fear, anxiety and anxiety-related disorders.

References

1 Berger, S. M. Conditioning through vicarious instigation. *Psychological Review* **69**, 450-466 (1962). https://doi.org/10.1037/h0046466

2 Vaughan, K. B. & Lanzetta, J. T. Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. *J Pers Soc Psychol* **38**, 909-923 (1980). https://doi.org/10.1037//0022-3514.38.6.909

3 Olsson, A. & Phelps, E. A. Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychol Sci* **15**, 822-828 (2004). https://doi.org/10.1111/j.0956-7976.2004.00762.x

4 Olsson, A., Nearing, K. I. & Phelps, E. A. Learning fears by observing others: the neural systems of social fear transmission. *Soc Cogn Affect Neurosci* **2**, 3-11 (2007). https://doi.org/10.1093/scan/nsm005

5 Haaker, J., Yi, J., Petrovic, P. & Olsson, A. Endogenous opioids regulate social threat learning in humans. *Nat Commun* **8**, 15495 (2017). https://doi.org/10.1038/ncomms15495

6 Olsson, A., Knapska, E. & Lindstrom, B. The neural and computational systems of social learning. *Nat Rev Neurosci* **21**, 197-212 (2020). https://doi.org/10.1038/s41583-020-0276-4

7 Lindstrom, B., Haaker, J. & Olsson, A. A common neural network differentially mediates direct and social fear learning. *Neuroimage* **167**, 121-129 (2018). https://doi.org/10.1016/j.neuroimage.2017.11.039

8 Marschner, A., Kalisch, R., Vervliet, B., Vansteenwegen, D. & Buchel, C. Dissociable roles for the hippocampus and the amygdala in human cued versus context fear conditioning. *J Neurosci* **28**, 9030-9036 (2008). https://doi.org/10.1523/JNEUROSCI.1651-08.2008

9 Quirk, G. J., Repa, C. & LeDoux, J. E. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely

behaving rat. *Neuron* **15**, 1029-1039 (1995). https://doi.org/10.1016/0896-6273(95)90092-6

10 Buchel, C., Holmes, A. P., Rees, G. & Friston, K. J. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage* **8**, 140-148 (1998). https://doi.org/10.1006/nimg.1998.0351

11 Buchel, C., Morris, J., Dolan, R. J. & Friston, K. J. Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* **20**, 947-957 (1998). https://doi.org/10.1016/s0896-6273(00)80476-6

12 Haaker, J., Golkar, A., Selbing, I. & Olsson, A. Assessment of social transmission of threats in humans using observational fear conditioning. *Nat Protoc* **12**, 1378-1386 (2017). https://doi.org/10.1038/nprot.2017.027

13 Peng, W., Huang, X., Liu, Y. & Cui, F. Predictability modulates the anticipation and perception of pain in both self and others. *Social Cognitive and Affective Neuroscience* **14**, 747-757 (2019). https://doi.org/10.1093/scan/nsz047

14 Schmitz, A. & Grillon, C. Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the NPU-threat test). *Nat Protoc* **7**, 527-532 (2012). https://doi.org/10.1038/nprot.2012.001

15 Alvarez, R. P., Chen, G., Bodurka, J., Kaplan, R. & Grillon, C. Phasic and sustained fear in humans elicits distinct patterns of brain activity. *Neuroimage* **55**, 389-400 (2011). https://doi.org/10.1016/j.neuroimage.2010.11.057

16 Naaz, F., Knight, L. K. & Depue, B. E. Explicit and Ambiguous Threat Processing: Functionally Dissociable Roles of the Amygdala and Bed Nucleus of the Stria Terminalis. *J Cogn Neurosci* **31**, 543-559 (2019). https://doi.org/10.1162/jocn_a_01369

17 Shankman, S. A. *et al.* Anterior insula responds to temporally unpredictable aversiveness: an fMRI study. *Neuroreport* **25**, 596-600 (2014). https://doi.org/10.1097/WNR.0000000000000144

18 Alvarez, R. P., Biggs, A., Chen, G., Pine, D. S. & Grillon, C. Contextual fear conditioning in humans: cortical-hippocampal and amygdala contributions. *J Neurosci* **28**, 6211-6219 (2008). https://doi.org/10.1523/JNEUROSCI.1246-08.2008

19 Figel, B. *et al.* Phasic amygdala and BNST activation during the anticipation of temporally unpredictable social observation in social anxiety disorder patients. *Neuroimage Clin* **22**, 101735 (2019). https://doi.org/10.1016/j.nicl.2019.101735

20 Geng, H. *et al.* Altered brain activation and connectivity during anticipation of uncertain threat in trait anxiety. *Hum Brain Mapp* **39**, 3898-3914 (2018). https://doi.org/10.1002/hbm.24219

21 Hur, J. *et al.* Anxiety and the Neurobiology of Temporally Uncertain Threat Anticipation. *J Neurosci* **40**, 7949-7964 (2020). https://doi.org/10.1523/JNEUROSCI.0704-20.2020

22 Radoman, M., Lieberman, L., Jimmy, J. & Gorka, S. M. Shared and unique neural circuitry underlying temporally unpredictable threat and reward processing. *Soc Cogn Affect Neurosci* **16**, 370-382 (2021). https://doi.org/10.1093/scan/nsab006

23 Herrmann, M. J. *et al.* Phasic and sustained brain responses in the amygdala and the bed nucleus of the stria terminalis during threat anticipation. *Hum Brain Mapp* **37**, 1091-1102 (2016). https://doi.org/10.1002/hbm.23088

24 Knobloch, S., Jansen, T. & Haaker, J. *nemo - fMRI*, <https://osf.io/5ny9u/> (2024).

25 Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. STAI manual for the Stait-Trait-Anxiety Inventory (Consulting Psycholo gist, Palo Alto, CA). (1970).

26 Vachon, D. D. & Lynam, D. R. Fixing the Problem With Empathy: Development and Validation of the Affective and Cognitive Measure of Empathy. *Assessment* **23**, 135-149 (2016). https://doi.org/10.1177/1073191114567941

27 Amunts, K. *et al.* Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl)* **210**, 343-352 (2005). https://doi.org/10.1007/s00429-005-0025-5

28 Lamm, C., Decety, J. & Singer, T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**, 2492-2502 (2011). https://doi.org/10.1016/j.neuroimage.2010.10.014

29 Jauniaux, J., Khatibi, A., Rainville, P. & Jackson, P. L. A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Soc Cogn Affect Neurosci* **14**, 789-813 (2019). https://doi.org/10.1093/scan/nsz055

30 Andreatta, M. *et al.* Initial and sustained brain responses to contextual conditioned anxiety in humans. *Cortex* **63**, 352-363 (2015). https://doi.org/10.1016/j.cortex.2014.09.014

31 Siminski, N. *et al.* BNST and amygdala activation to threat: Effects of temporal predictability and threat mode. *Behav Brain Res* **396**, 112883 (2021). https://doi.org/10.1016/j.bbr.2020.112883

32 Wheelock, M. D. *et al.* Threat-related learning relies on distinct dorsal prefrontal cortex network connectivity. *Neuroimage* **102 Pt 2**, 904-912 (2014). https://doi.org/10.1016/j.neuroimage.2014.08.005

33 Krabbe, S., Grundemann, J. & Luthi, A. Amygdala Inhibitory Circuits Regulate Associative Fear Conditioning. *Biol Psychiatry* **83**, 800-809 (2018). https://doi.org/10.1016/j.biopsych.2017.10.006

34 Koo, J. W., Han, J. S. & Kim, J. J. Selective neurotoxic lesions of basolateral and central nuclei of the amygdala produce differential effects on fear conditioning. *J Neurosci* **24**, 7654-7662 (2004). https://doi.org/10.1523/JNEUROSCI.1644-04.2004

35 Cornwell, B. R. *et al.* A Shared Threat-Anticipation Circuit Is Dynamically Engaged at Different Moments by Certain and Uncertain Threat. *J Neurosci* **45** (2025). https://doi.org/10.1523/JNEUROSCI.2113-24.2025

36 Robinson, J. L. *et al.* Neurofunctional topography of the human hippocampus. *Hum Brain Mapp* **36**, 5018-5037 (2015). https://doi.org/10.1002/hbm.22987

37 Cooper, S. E., Hennings, A. C., Bibb, S. A., Lewis-Peacock, J. A. & Dunsmoor, J. E. Semantic structures facilitate threat memory integration throughout the medial temporal lobe and medial prefrontal cortex. *Curr Biol* **34**, 3522-3536 e3525 (2024). https://doi.org/10.1016/j.cub.2024.06.071

38 Quirk, G. J., Armony, J. L. & LeDoux, J. E. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19**, 613-624 (1997). https://doi.org/10.1016/s0896-6273(00)80375-x

39 Strange, B. A., Fletcher, P. C., Henson, R. N., Friston, K. J. & Dolan, R. J. Segregating the functions of human hippocampus. *Proc Natl Acad Sci U S A* **96**, 4034-4039 (1999). https://doi.org/10.1073/pnas.96.7.4034

40 Rescorla, R. A., Wagner, A. R., Black, A. H. & Prokasy, W. F. in *Classical conditioning II: Current research and theory* 64-99 (Appleton-Century-Crofts, 1972).

41 Duvarci, S., Popa, D. & Pare, D. Central amygdala activity during fear conditioning. *J Neurosci* **31**, 289-294 (2011). https://doi.org/10.1523/JNEUROSCI.4985-10.2011

42 Wesarg-Menzel, C., Gallistl, M., Niconchuk, M. & Engert, V. Adding to the burden: the tendency to resonate with others' stress is linked to higher PTSD symptom severity in individuals with war-related trauma. *Transl Psychiatry* **15**, 331 (2025). https://doi.org/10.1038/s41398-025-03548-4

43 Chen, P.-H. A. *et al.* Socially transmitted placebo effects. *Nature Human Behaviour* **3**, 1295-1305 (2019). https://doi.org/10.1038/s41562-019-0749-5

44 Cooper, S. E. *et al.* A meta-analysis of conditioned fear generalization in anxiety-related disorders. *Neuropsychopharmacology* **47**, 1652-1661 (2022). https://doi.org/10.1038/s41386-022-01332-2

45 Cooper, S. E. & Dunsmoor, J. E. Fear conditioning and extinction in obsessive-compulsive disorder: A systematic review. *Neurosci Biobehav Rev* **129**, 75-94 (2021). https://doi.org/10.1016/j.neubiorev.2021.07.026

Figure legends

Figure 1: Experimental Design.

a) shows three extracted frames from the P, U and N stimulus material respectively. To increase visibility, the cue position is marked with a red circle and the painful stimulation is marked with a red bolt. These markups were not visible for the participants. In P, the demonstrator is visible at rest while the slider passes the grey parts of the rectangle before and after the cue. The demonstrator is in pain during the cue (purple part of the rectangle). In U, the demonstrator is visible at rest while the slider passes the grey part of the rectangle before the cue and during cue (yellow part of the rectangle). The demonstrator is in pain during the passing of the grey part of the rectangle after the cue. In N, the demonstrator is at rest at all slider positions. b) gives an overview of the condition design. Each condition is revealed at stimulus onset, as only one slider starts moving. Unpredictable and predictable threat, thus, can be anticipated at each condition onset. The crosshatched red lines mark the typical pain position in P and a possible pain position in U. The crosshatched lines were not visible for the participants.

Figure 2: Behavioral results.

Figure 2 depicts bar plots with superimposed scatter element. Bars represent the mean expectation of pain during context-onset (2.5s before cue onset) and during cue (2.5s following cue onset).

Figure 3: Observation Phase Pain.

a)-f) depict the contrast P & U > N.

a) shows a 80x80 voxel zoom in coronary view of the left basolateral amygdala cluster. b) depicts a full volume sagittal view of the same cluster. c) shows diagrams of the peak-voxel ß-estimates of the left amygdala cluster. The respective cluster is highlighted with a red circle in a) and b).

d) shows a 80x80 voxel zoom in coronary view of the left insula cluster. e) depicts a full volume sagittal view of the same cluster. f) shows diagrams of the peak-voxel ß-estimates of the left insula cluster. The respective cluster is highlighted with a red circle in d) and e).

g) – i) depict the contrast P > U.

g) shows a 80x80 voxel zoom in coronary view of the right hippocampus cluster. g) depicts a full volume sagittal view of the same cluster. i) shows diagrams of the peak-voxel ß-estimates of the right hippocampus cluster. The respective cluster is highlighted with a red circle in g) and h).

Figure 4: Cue Prominence

a)-c) depict the cue-prominence contrast (context: U > P, cue: P > U) of the observation phase. Note that this cluster only trends significance in FWE correction (p = 0.056).
a) shows a 80x80 voxel zoom in coronary view of the left hippocampus cluster. b) depicts a full volume sagittal view of the same cluster. c) shows diagrams of the peak-voxel ß-estimates of the left hippocampus cluster. The respective cluster is highlighted with a red circle in a) and b).
d) – e) depict the cue-prominence contrast (context: U > P, cue: P > U) of the expression phase. d) shows a 80x80 voxel zoom in coronary view of the right insula cluster. e) depicts a full volume sagittal view of the same cluster. f) shows diagrams of the peak-voxel ß-estimates of the right insula cluster. The respective cluster is highlighted with a red circle in d) and e).
Note: light shaded bars in both plots of the beta estimates c) and f) represent the estimates for the respective context-onsets, dark shaded bars represent the estimates for the respective cue.

Figure 5: Observation Phase: Cue

a)-c) depict the contrast P > N of the cue.
a) shows a 80x80 voxel zoom in coronary view of the left hippocampus cluster. b) depicts a full volume sagittal view of the same cluster. c) shows diagrams of the peak-voxel ß-estimates of the left hippocampus cluster. The respective cluster is highlighted with a red circle in a) and b).

Figure 6: Observation Phase: Context

a)-c) depict the contrast U > N of the context-onset.
a) shows a 80x80 voxel zoom in coronary view of the left insula cluster. b) depicts a full volume sagittal view of the same cluster. c) shows diagrams of the peak-voxel ß-estimates of the left insula cluster. The respective cluster is highlighted with a red circle in a) and b).
d) – e) depict the contrast U > P of the sustained context activity.
d) shows a 80x80 voxel zoom in coronary view of the right basolateral amygdala cluster. e) depicts a full volume sagittal view of the same cluster. f) shows diagrams of the peak-voxel ß-estimates of the right amygdala cluster. The respective cluster is highlighted with a red circle in d) and e).

Figure 7: tMod cue-prominence.

a)-c) depict the linearly decreasing interaction contrast of the cue-prominence in the observation phase.
a) shows a 80x80 voxel zoom in coronary view of the right insula cluster. b) depicts a full volume sagittal view of the same cluster. c) shows diagrams of the peak-voxel ß-estimates of the right insula cluster. The respective cluster is highlighted with a red circle in a) and b).
d)-e) depict the linearly decreasing interaction contrast of the cue-prominence in the expression phase.
d) shows a 80x80 voxel zoom in coronary view of the right centromedial and basolateral amygdala cluster. e) depicts a full volume sagittal view of the same cluster. f) shows diagrams of the peak-voxel ß-estimates of the right amygdala cluster. The respective cluster is highlighted with a red circle in d) and e).
Note: light shaded bars in both plots of the beta estimates c) and f) represent the estimates for the respective context-onsets, dark shaded bars represent the estimates for the respective cue.

Figure 8: tMod cue and context

a)-c) depict the linearly decreasing contrast P > U of the cue in the observation phase.
a) shows a 80x80 voxel zoom in coronary view of the bilateral centromedial amygdala clusters. b) depicts a full volume sagittal view of the same clusters. c) shows diagrams of the peak-voxel ß-estimates of both amygdala clusters. The respective clusters are highlighted with red circles in a) and b).
d)-e) depict the linearly decreasing contrast P > U of the cue in the expression phase.
d) shows a 80x80 voxel zoom in coronary view of the right centromedial amygdala cluster. e) depicts a full volume sagittal view of the same cluster. f) shows diagrams of the peak-voxel ß-estimates of the right amygdala cluster. The respective cluster is highlighted with a red circle in d) and e).
g) – i) depict the linearly decreasing contrast P > U of the sustained context activity in the expression phase.
g) shows a 80x80 voxel zoom in coronary view of the left hippocampus cluster. g) depicts a full volume sagittal view of the same cluster. i) shows diagrams of the peak-voxel ß-estimates of the left hippocampus cluster. The respective cluster is highlighted with a red circle in g) and h).

Tables and table legends

Table 1. Sample characteristics Experiment 1

Characteristics	Descriptives
Age (mean ± SD)	25.28 ± 3.65
Self-identified Gender	
Male, n	6
Female, n	14
Diverse, n	0
Highest Degree in Education	
Secondary School Diploma, n	2
High School Diploma, n	8
University Degree, n	8
Medical Background	
Yes, n	7
No, n	10

Table 1 legend. Table 1 reports the sample characteristics of the participants of Experiment 1. We report age (mean ± standard deviation), self-identified gender, education and the binary variable medical background. Participants reported a medical background when they worked or studied in medicine or related fields like e.g. nursing.

Table 2. Stimuli of Experiment 1 and 2: Descriptives of measures characterizing the video stimuli.

	Experiment 1			Experiment 2		
Measure [seconds]	**P**	**U**	**N**	**P**	**U**	**N**
Cue length (mean ± SD)	7.8 ± 1.6	8.6 ± 2.4	8.4 ± 3.5	7.5 ± 1.5	8.5 ± 2.7	7.3 ± 2.0
Pain length (mean ± SD)	6.7 ± 1.2	7.0 ± 1.3		6.4 ± 1.2	6.8 ± 1.4	
Overlap (mean ± SD)	5.1 ± 1.6	3.6 ± 2.7		4.9 ± 1.6	3.7 ± 2.6	

Table 2 legend. Table 2 reports a description of the video stimuli used in the experiment. We report mean and standard deviation [SD] of the length of the cue, the length of the visible pain and the overlap of cue and pain. Measures are given in seconds.

Table 3. Sample characteristics Experiment 2

Characteristics	Descriptives
Age (mean ± SD)	27.3 ± 6.14
Self-identified Gender	
Male, n	13
Female, n	9
Diverse, n	1
Highest Degree in Education	
Secondary School Diploma, n	3
High School Diploma, n	12
University Degree, n	8
Medical Background	
Yes, n	5
No, n	18
Questionnaires	
STAI – State (mean ± SD)	33.6 ± 7.9
STAI – Trait (mean ± SD)	32.2 ± 7.3
ACME – Cognitive Empathy (mean ± SD)	27.2 ± 8.4
ACME – Affective Resonance (mean ± SD)	20.7 ± 8.9
ACME – Affective Dissonance (mean ± SD)	20.8 ± 11.4

Table 3 legend. Table 3 reports the sample characteristics of the participants of Experiment 2. We report age (mean ± standard deviation), self identified gender, education and the binary variable medical background. Participants reported a medical background when they worked or studied in medicine or related fields like e.g. nursing. Furthermore, we report the computed variables from the STAI and ACME questionnaires. Mean and standard deviation are reported.

Table 4. Experiment 1: Main effect repeated measurement ANOVA. Threat Expectation

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Within Subjects Effects						
Condition	8.592	2	4.296	15.175	<.001	0.444
Residuals	10.758	38	0.283			
Cue/Context-onset	21.931	1	21.931	62.033	< .001	0.766
Residuals	6.717	19	0.354			
Condition X Cue/Context-onset	2.455	2	1.228	9.383	< .001	0.331
Residuals	4.972	38	0.131			

Table 4 legend. Table 4 reports the detailed results of the behavioral effect (threat expectation) rmANOVA (3x2: Condition: P, U, N; Timepoint: cue, context-onset) of Experiment 1. We report sum of squares, degrees of freedom, mean square, F-values, p and an estimation of the effect sizes (partial η^2) for the main effects of condition and timepoint and the interaction effect.

Table 5. Experiment 1: Paired-sample t-tests. Learned cue-prominence after observation phase.

Variable 1	**Variable 2**	**df**	**T**	**p**	$\mathbf{p_{Holm}}$	**Cohen's d**
Cue-Prominence P	Cue-Prominence N	19	3.994	.00078	.00234	0.893
Cue-Prominence U	Cue-Prominence N	19	2.449	.024	.048	0.548
Cue-Prominence P	Cue-Prominence U	19	1.997	.06	.06	0.447

Table 5 legend. Table 5 reports the posthoc paired-sample t-tests conducted to analyze the significant interaction effect of the rmANOVA in Experiment 1. We compare the learned cue-prominence between all three conditions. Degrees of freedom, T, p, p corrected for multiple comparisons (Bonferroni-Holm) and an estimation of effect size (Cohen's d) are reported.

Table 6. Experiment 1: Behavioral outcome measures

Variable	Descriptives
Cue-Prominence P (mean ± SD)	117.4 ± 63.4
Cue-Prominence U (mean ± SD)	91.1 ± 65.1
Cue-Prominence N (mean ± SD)	48.0 ± 63.6
Pain-rating P (mean ± SD)	51.1 ± 19.9
Pain-rating U (mean ± SD)	49.3 ± 21.3

Table 6 legend. Table 6 reports the descriptive results of Experiment 1. We report the calculated cue-prominence for each condition (P, U, N) and the mean pain-ratings for P and U. Mean values and standard deviations are reported.

Table 7. Experiment 2: Main effect repeated measurement ANOVA. Threat Expectation

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Within Subjects Effects						
Condition	9.729[a]	2[a]	4.865[a]	22.642[a]	<.001[a]	0.507
Residuals	9.453	44	0.215			
Cue/Context-onset	9.926	1	9.926	32.291	< .001	0.595
Residuals	6.762	22	0.307			
Condition X Cue/Pr-Cue Context	2.022	2	1.011	13.158	< .001	0.374
Residuals	3.381	44	0.077			

Table 7 legend. Table 7 reports the detailed results of the behavioral effect (threat expectation) rmANOVA (3x2: Condition: P, U, N; Timepoint: cue, context-onset) of Experiment 2. We report sum of squares, degrees of freedom, mean square, F-values, p and an estimation of the effect sizes (partial η^2). □ points at the violation of the assumption of sphericity (Mauchly's test, $p < .05$).

Table 8. Experiment 2: Behavioral outcome measures

Variable	Descriptives
Initial Cue-Prominence P (mean ± SD)	55.7 ± 79.6
Initial Cue-Prominence U (mean ± SD)	31.6 ± 55.9
Initial Cue-Prominence N (mean ± SD)	48.2 ± 62.6
Cue-Prominence P (mean ± SD)	82.8 ± 60.8
Cue-Prominence U (mean ± SD)	54.7 ± 56.8
Cue-Prominence N (mean ± SD)	23.5 ± 47.9
Pain-rating P (mean ± SD)	62.6 ± 20.2
Pain-rating U (mean ± SD)	66.5 ± 21.9

Table 8 legend. Table 8 reports the descriptive results of Experiment 1. We report the calculated cue-prominence for each condition (P, U, N) before (initial cue-prominence) and after learning and the mean pain-ratings for P and U. Mean values and standard deviations are reported.

Table 9. Experiment 2: Paired-sample t-tests. Learned cue-prominence after observation phase.

Variable 1	Variable 2	df	T	p	p_{Holm}	Cohen's d
Learned cue-prominence P	Learned cue-prominence N	22	4.443	.0002	.0006	.926
Learned cue-prominence U	Learned cue-prominence N	22	2.814	.01	.02	.587
Learned cue-prominence P	Learned cue-prominence U	22	2.808	.01	.01	.586

Table 9 legend. Table 9 reports the posthoc paired-sample t-tests conducted to analyze the significant interaction effect of the rmANOVA in Experiment 2. We compare the learned cue-prominence between all three conditions. Degrees of freedom, T, p, p corrected for multiple comparisons (Bonferroni-Holm) and an estimation of effect size (Cohen's d) are reported.

Table 10. Experiment 2: Paired-sample t-tests. Initial cue-prominence before observation phase.

Variable 1	Variable 2	df	T	p	p_{Holm}
Initial cue-prominence P	Initial cue-prominence N	22	0.943	.356	.356
Initial cue-prominence U	Initial cue-prominence N	22	-2.443	.023	.069
Initial cue-prominence P	Initial cue-prominence U	22	2.066	.052	.104

Table 10 legend. Table 10 reports the paired-sample t-tests conducted to analyze the cue-prominence previous to the observational learning phase in Experiment 2. We compare the learned cue-prominence between all three conditions. Degrees of freedom, T, p and p corrected for multiple comparisons (Bonferroni-Holm) are reported.

Figures

Figure 1

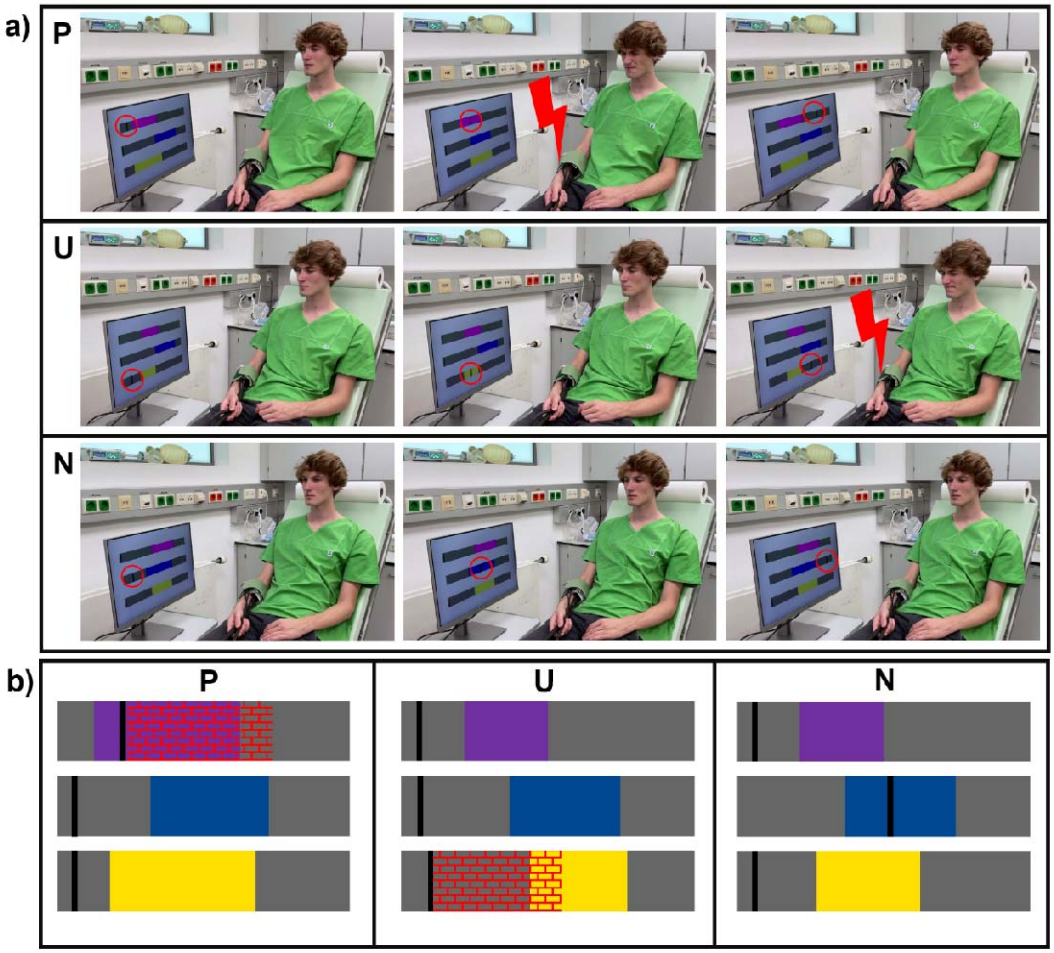


Figure 2

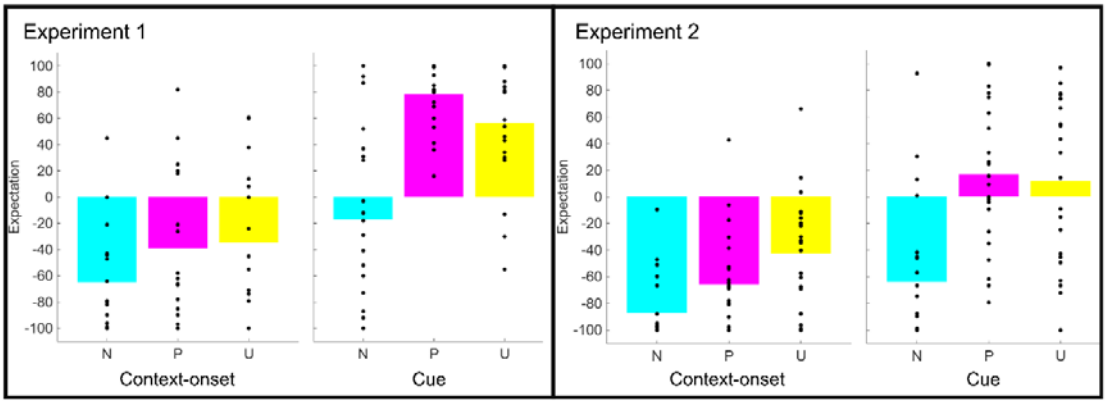

Experiment 1
Experiment 2
Expectation
100
80
60
40
20
0
-20
-40
-60
-80
-100
N
P
U
Context-onset
Cue

Figure 3

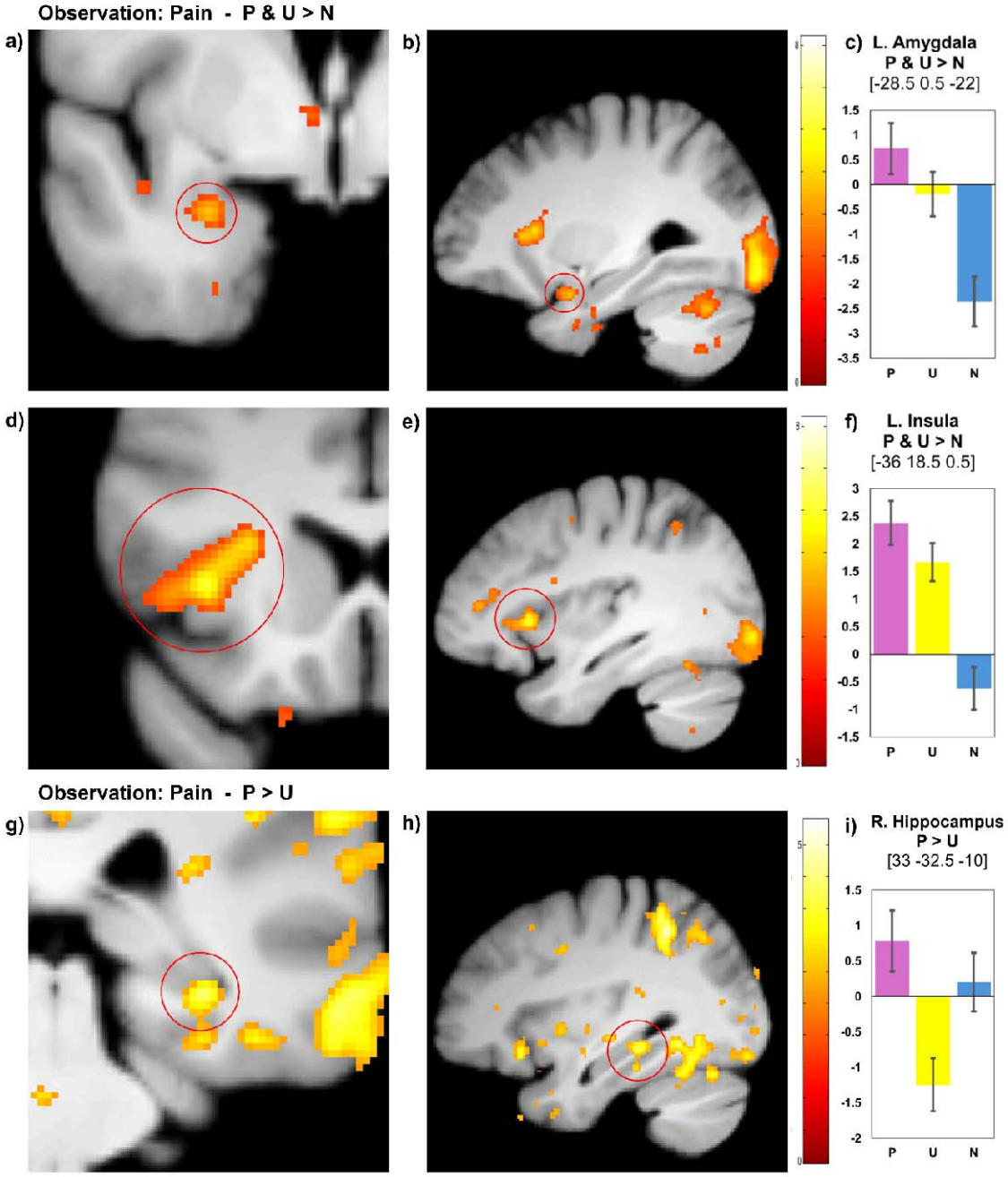

Observation: Pain - P & U > N
a)
b)
c) L. Amygdala
P & U > N
[-28.5 0.5 -22]
d)
e)
f) L. Insula
P & U > N
[-36 18.5 0.5]
Observation: Pain - P > U
g)
h)
i) R. Hippocampus
P > U
[33 -32.5 -10]
P
U
N

Figure 4

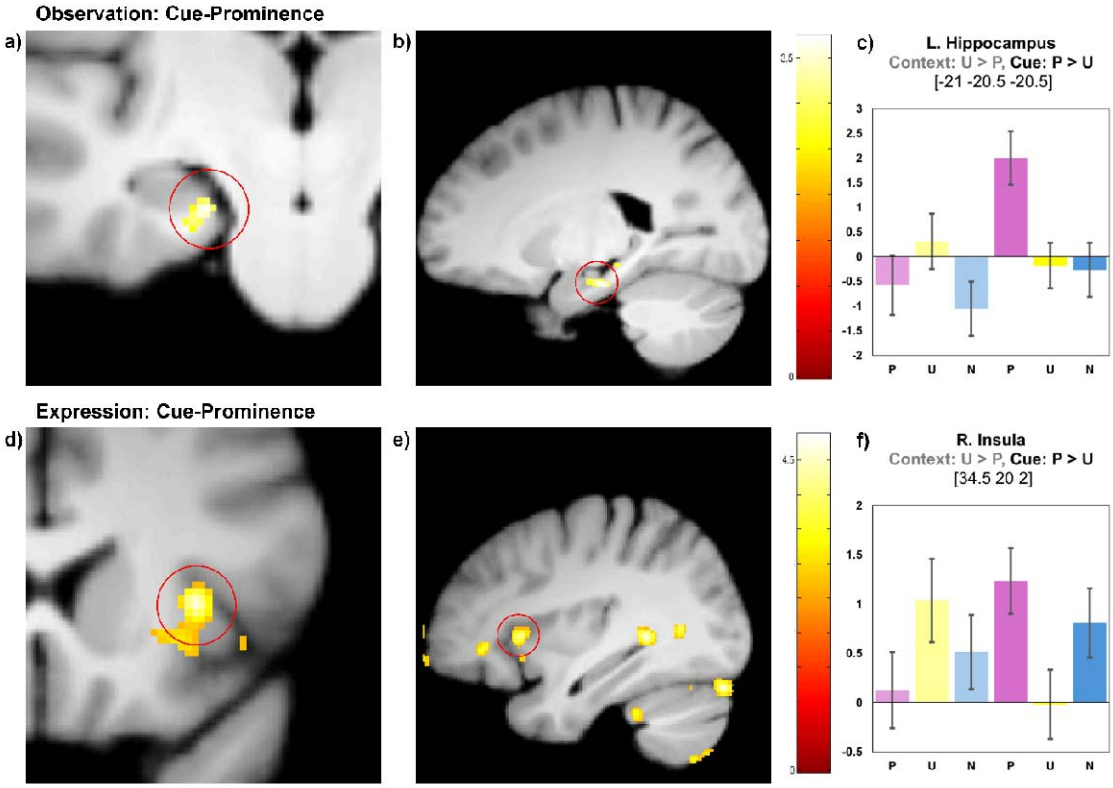

Observation: Cue-Prominence
a)
b)
c)
L. Hippocampus
Context: U > P, Cue: P > U
[-21 -20.5 -20.5]
P U N P U N
Expression: Cue-Prominence
d)
e)
f)
R. Insula
Context: U > P, Cue: P > U
[34.5 20 2]
P U N P U N

Figure 5

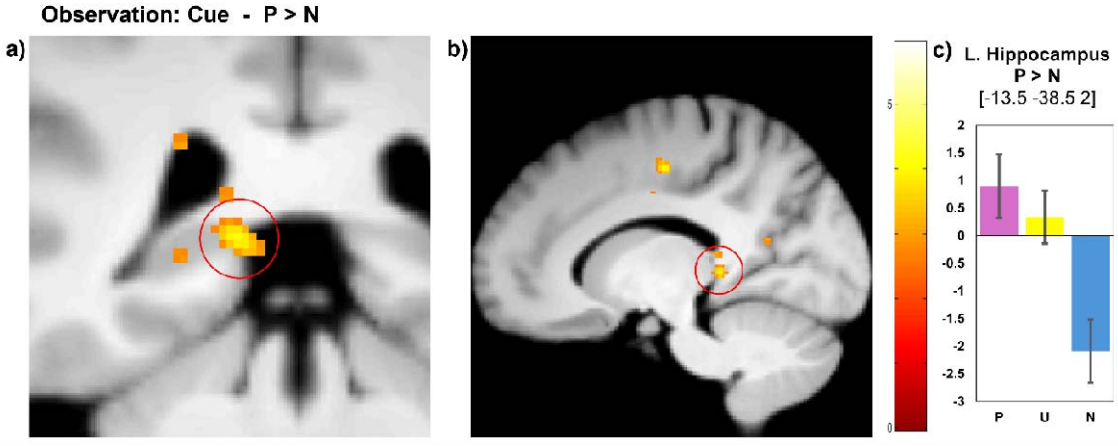

Observation: Cue - P > N
a)
b)
c) L. Hippocampus
P > N
[-13.5 -38.5 2]
5
0
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
P
U
N

Figure 6

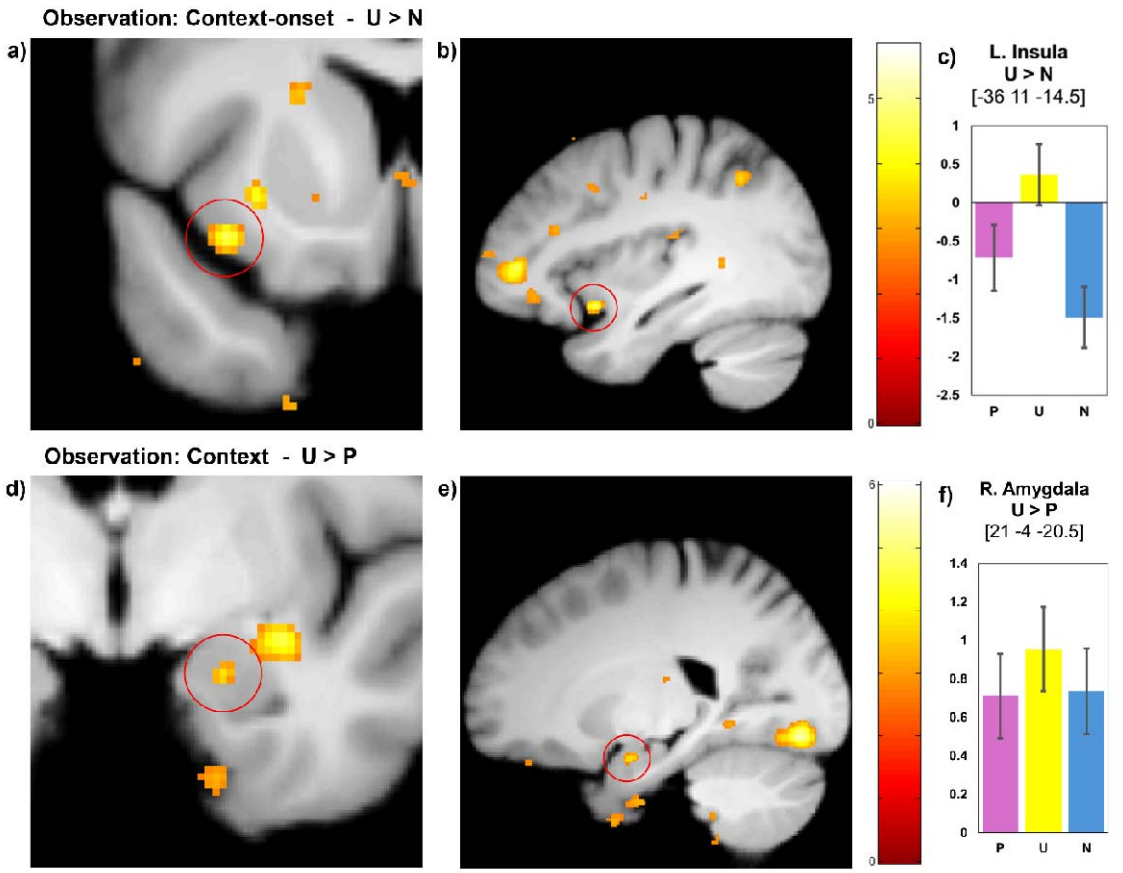

Observation: Context-onset - U > N
a)
b)
c)
L. Insula
U > N
[-36 11 -14.5]
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
P
U
N
5
0
Observation: Context - U > P
d)
e)
f)
R. Amygdala
U > P
[21 -4 -20.5]
1.4
1.2
1
0.8
0.6
0.4
0.2
0
P
U
N
6
0

Figure 7

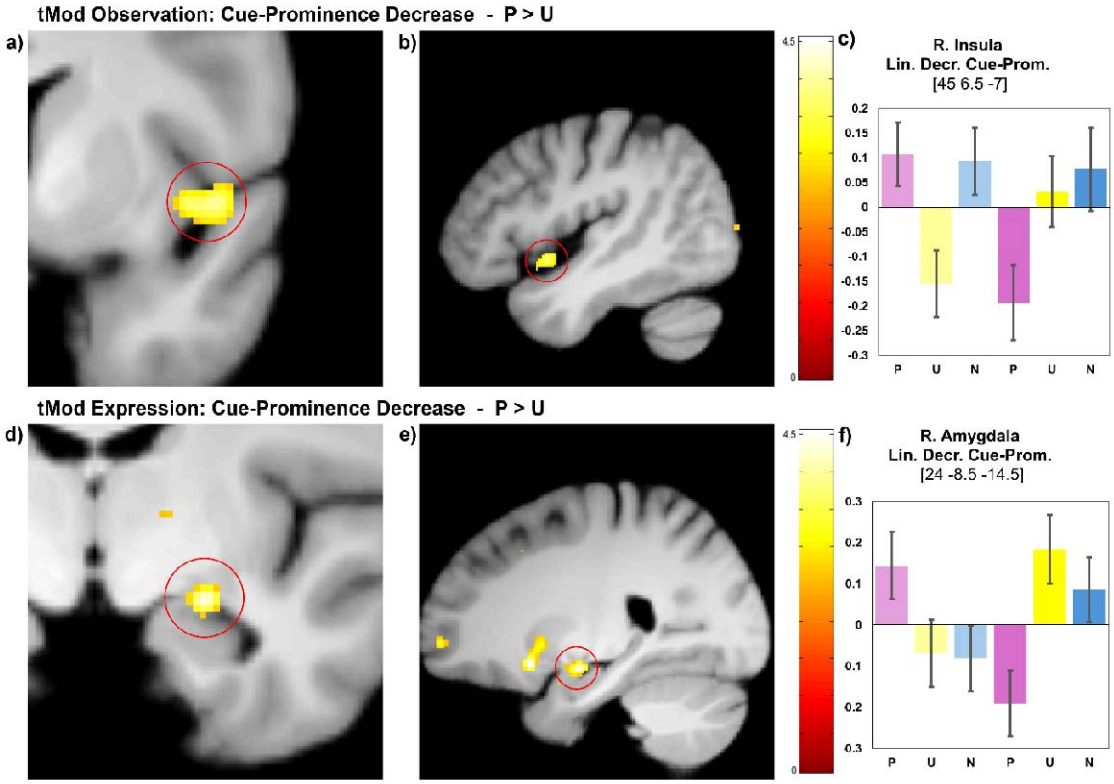
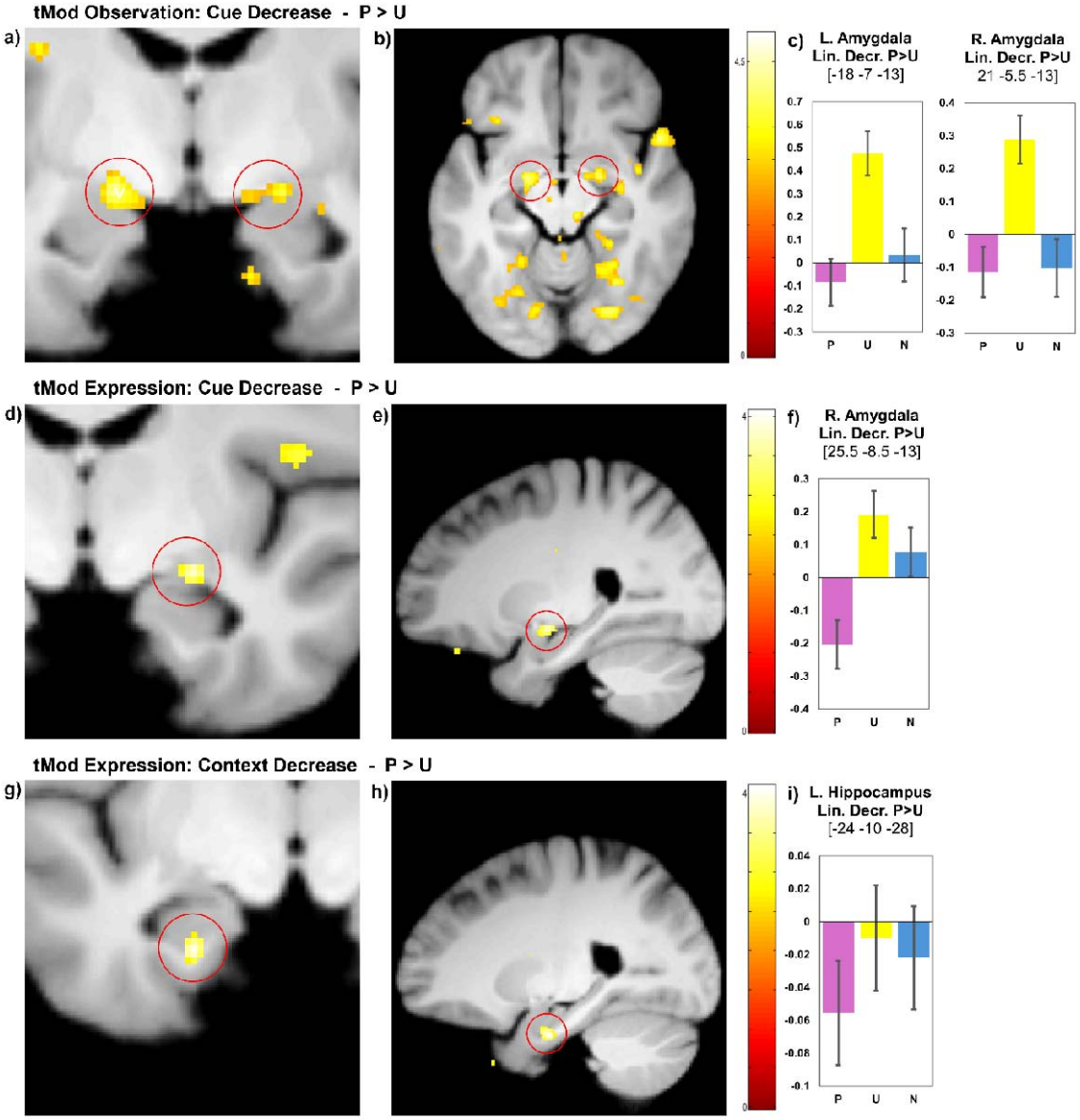

tMod Observation: Cue-Prominence Decrease - P > U
a)
b)
c)
R. Insula
Lin. Decr. Cue-Prom.
[45 6.5 -7]
P U N P U N
tMod Expression: Cue-Prominence Decrease - P > U
d)
e)
f)
R. Amygdala
Lin. Decr. Cue-Prom.
[24 -8.5 -14.5]
P U N P U N

Figure 8


tMod Observation: Cue Decrease - P > U
a)
b)
c)
L. Amygdala
Lin. Decr. P>U
[-18 -7 -13]
R. Amygdala
Lin. Decr. P>U
21 -5.5 -13]
P U N
tMod Expression: Cue Decrease - P > U
d)
e)
f)
R. Amygdala
Lin. Decr. P>U
[25.5 -8.5 -13]
P U N
tMod Expression: Context Decrease - P > U
g)
h)
i)
L. Hippocampus
Lin. Decr. P>U
[-24 -10 -28]
P U N