Social Cognitive and Affective Neuroscience, 2018, 189–203

doi: 10.1093/scan/nsy103
Advance Access Publication Date: 27 November 2018
Original article

Learning to see the threat: temporal dynamics of ERPs of motivated attention in fear conditioning

Diana S. Ferreira de Sá,[1] Tanja Michael,[1] Frank H. Wilhelm,[2] and Peter Peyk[3]

[1]Division of Clinical Psychology and Psychotherapy, Department of Psychology, Saarland University, D-66123 Saarbrücken, Germany, [2]Division of Clinical Psychology, Psychotherapy, and Health Psychology, Department of Psychology, University of Salzburg, 5020 Salzburg, Austria and [3]Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, 8091 Zurich, Switzerland

Correspondence should be addressed to Peter Peyk, Haldenbachstrasse 16/18, 8006 Zurich, Switzerland. E-mail: peter.peyk@usz.ch.

Abstract

Social threat detection is important in everyday life. Studies of cortical activity have shown that event-related potentials (ERPs) of motivated attention are modulated during fear conditioning. The time course of motivated attention in learning and extinction of fear is, however, still largely unknown. We aimed to study temporal dynamics of learning processes in classical fear conditioning to social cues (neutral faces) by selecting an experimental setup that produces large effects on well-studied ERP components (early posterior negativity, EPN; late positive potential, LPP; stimulus preceding negativity, SPN) and then exploring small consecutive groups of trials. EPN, LPP, and SPN markedly and quickly increased during the acquisition phase in response to the CS^+ but not the CS^-. These changes were visible even at high temporal resolution and vanished completely during extinction. Moreover, some evidence was found for component differences in extinction learning, with differences between CS+ and CS− extinguishing faster for late as compared to early ERP components. Results demonstrate that fear learning to social cues is a very fast and highly plastic process and conceptually different ERPs of motivated attention are sensitive to these changes at high temporal resolution, pointing to specific neurocognitive and affective processes of social fear learning.

Key words: affective neuroscience; fear conditioning; motivated attention; social threat; event-related potentials; temporal dynamics

Introduction

Human beings are subjected every day to an intense flow of stimuli, with many of them carrying motivationally relevant information. Detection of these cues and the ability to prioritize their processing plays an important role in adaptation and survival. In particular, emotional stimuli capture attention, fitting a model of 'motivated attention' (Lang *et al.*, 1997). Studies of motivated attention have identified various markers in brain activity that reflect the enhanced processing of such cues (e.g. Keil *et al.*, 2001; Bradley *et al.*, 2003; Moratti *et al.*, 2004; Peyk *et al.*, 2008; Brunia *et al.*, 2011).

Clinical studies suggest that cortical processing of emotional stimuli is altered in several mental disorders (Flor *et al.*, 2002; Felmingham *et al.*, 2003; Ehlers *et al.*, 2006; Michalowski *et al.*, 2009; Van Strien *et al.*, 2014), raising the question of how this altered processing is first acquired. Altered fear conditioning processes have been suggested as important factors in the etiology and maintenance of anxiety disorders and posttraumatic stress disorder (Dye and Blundell, 1997; Lang *et al.*, 2000; Blechert *et al.*, 2007; Michael *et al.*, 2007; Wessa and Flor, 2007). Fear associative learning can be operationalized by pairing a neutral stimulus (conditioned stimulus, CS) with an

Received: 1 December 2017; **Revised:** 13 November 2018; **Accepted:** 21 November 2018

aversive one (unconditioned stimulus, US). In the acquisition phase, participants learn that the neutral CS signals the unpleasant US, leading to the development of a preparatory fear response (conditioned response, CR). This CR will then precede the organism's unconditioned response to the aversive US. In an extinction phase, the same CS is repeatedly presented alone (no US), resulting in a decrease of the CR, in what is hypothesized to be the result of an additional learning process (Bouton, 2004).

Social anxiety disorder is the most common anxiety disorder (Stein and Stein, 2008) and, even at subclinical levels, it is associated with higher probability of comorbid disorders and greater impairment in different life domains (Fehm *et al*., 2008). In modern society, social interactions are an important source of stress, but also crucial for daily life and difficult to avoid. Social interactions perceived as threatening can be a source of continued fear and chronic stress (Tost *et al*., 2015). Wiggert *et al*., (2016) recently demonstrated that social learning by Pavlovian conditioning was maintained even after 1 month and 1 year follow-ups.

Over the past years, approaches such as the Research Domain Criteria framework, launched by the National Institute of Mental Health, have emphasized that a profound understanding of basic neurobiological functional dimensions underlying human behavior is essential to foster knowledge on psychological disorders (Insel, 2014). It is therefore important to study the plasticity and time dynamics of motivated attention during social fear conditioning to allow a deeper comprehension of this process, which is disrupted in anxiety disorders.

Specific event-related potential (ERP) components like the early posterior negativity (EPN), the late positive potential (LPP) and the stimulus-preceding negativity (SPN) are prominent examples of markers of motivated attention (for a review, see Schupp *et al*., 2006a). The EPN is a relative negative shift in ERP waves over parieto-temporal-occipital sites ~150–300 ms after stimulus onset (Schupp *et al*., 2006a). It is assumed that the EPN represents 'natural selective attention' and is closely related to the activation of an evolutionary appetitive/aversive motivational system (Schupp *et al*., 2006a). The LPP is a positive shift in ERP waves over parieto-central sites ~400–600 ms after stimulus onset (Schupp *et al*., 2006a). Modulation of LPP amplitudes is thought to be related to the representation of the emotional cue in working memory (Schupp *et al*., 2006a). The SPN is a slow brain potential that occurs some seconds before a stimulus or feedback presentation. In terms of morphology, the SPN shows a right hemisphere dominance and enhanced negativity over parietal sites (for a review, see Brunia *et al*., 2011). Previous research has highlighted that the SPN reflects not only perceptual anticipation, but also emotional anticipation, which can be manipulated by motivational value (Kotani *et al*., 2003; Ohgami *et al*., 2004; Poli *et al*., 2007). SPN activity is mainly observed in anticipation to emotional cues, like fear-inducing stimuli (Bocker *et al*., 2001; Baas *et al*., 2002).

Electro-cortical evidence regarding threat learning and extinction is, however, still scarce and inconsistent (for a review, see Miskovic and Keil, 2012). Although there seems to be a consensus on the fact that differential CS processing can be seen in electro-cortical responses in general, the exact timing and topographical properties of these effects and their respective functional meaning remain unclear. Previous studies with socially relevant fear conditioning paradigms (e.g. using faces as CS) have investigated effects within the first second of stimulus processing, starting at around 50 ms after stimulus onset (e.g. Steinberg *et al*., 2012; Mueller and Pizzagalli, 2016). They have extended the effects on the P1 (Pizzagalli *et al*. 2003), the P2, the P3 (Rothemund *et al*., 2012) and LPP (Panitz *et al*., 2015) to slow cortical potentials, lasting up to several seconds (Flor *et al*., 1996). One reason for the diversity of the reported effects may be the fact that they are likely to comprise a multitude of different neurocognitive processes involved in fear learning and extinction. It is therefore not surprising that effects of altered or amplified processing have been reported to involve modality-dependent sensory cortices (e.g. auditory cortex: Dolan *et al*., 2006; Brockelmann *et al*., 2011; Kluge *et al*., 2011; Steinberg *et al*., 2012; visual cortex: Dolan *et al*., 2006; Keil *et al*., 2007; Liu *et al*., 2012; Steinberg *et al*., 2012) or parietal attention-related areas (Moratti *et al*., 2006; Liu *et al*., 2012; Rothemund *et al*., 2012), whereas evaluative, executive and memory processes seem to involve frontal, temporal and subcortical structures (Pizzagalli *et al*., 2003; Santesso *et al*., 2008; Mueller *et al*., 2014; Mueller and Pizzagalli, 2016; Sperl *et al*., 2018).

The diversity of reported effects observable in the literature may also be due to the low-to-medium spatial resolution of the Electroencephalogram (EEG) and magneto-encephalogram (MEG), which makes the interpretation of EEG or MEG effects a challenge. In addition, ever more complex algorithms available for pre- and post-processing, source localization and time–frequency analysis have introduced many new degrees of freedom that offer exciting new possibilities, but can also impair comparability and replication of findings. Up to date, not many studies have focused on the time course of attention in the learning and extinction of fear most likely due to difficulties in trial-by-trial analysis (Miskovic and Keil, 2012). To our knowledge, specifically, the temporal emergence of the EPN, LPP and SPN to affective value has not yet been explored.

The present study aimed to start filling this gap by carefully selecting an experimental setup of classical fear conditioning that produces very robust conditioning effects (specific changes on the CS across experimental phases: habituation, acquisition, extinction) and then systematically subdividing the trials available in each experimental phase into smaller subsets of trials to assess the trajectories of the learning curves with increasing temporal resolution.

Unless ERP components are estimated using more sophisticated modeling techniques (for an overview, see Blankertz *et al*., 2011), the calculation of standard evoked potentials relies on the averaging of a set of trials. Using this approach, the analysis of temporal 'changes' of evoked potentials can only be performed on consecutive sets of trials. The temporal resolution of such an analysis is a function of the exact number of trials within a set, and this number largely depends on the signal-to-noise ratio of the EEG signal. Very early components with small amplitudes (such as auditory brain stem potentials for instance) require a large number (several hundreds) of trials to be detected in the ERP. Large amplitude components such as slow waves may only require a handful of trials and may even be visible in the raw EEG signal. Additionally, the signal-to-noise ratio depends on the measurement error and thus the hardware used for recording of the EEG signal and the environment in which the data was collected. The current study's signal-to-noise ratio turned out to be excellent, allowing us to meaningfully analyze the selected ERP components of motivated attention (EPN, LPP and SPN) with a high temporal resolution (corresponding to small sets of trials) and to see whether and how their effect trajectory differs across acquisition and extinction phases.

We expected that during the acquisition phase, motivated attention, as measured by EPN, LPP and SPN, would be selectively increased for CS^+ (human face predicting the US) relative to the CS^- (human face never paired with the US), due to the associ-

ation of a negative emotional salience to the previously neutral face. We additionally expected that during the extinction phase, differences in selective and anticipatory attention between the CS^+ and the CS^- would disappear because of restored emotional salience to neutral levels. We expected these effects to fade with the number of trials per set decreasing below around 20, but without any precise expectation about the exact position of this threshold. Based on the comparably low number of US reinforcements that have been necessary for conditioning to occur in other studies (e.g. Steinberg *et al*., 2012), we expected learning to occur quickly in a matter of a few trials. Concerning the extinction, we expected the disappearance to take somewhat longer than the acquisition, based on previous exploratory findings, where after a short extinction the conditioning effects for the LPP were still visible (Bacigalupo and Luck, 2018).

Material and methods

Participants

Participants were 24 healthy right-handed students (12 females; mean age, 24.1 years; s.d., 4.8; range, 18–38) from the University of Basel, Switzerland. The psychological and physical health statuses were assessed by a screening questionnaire. The exclusion criteria were current health complaints, abuse of illicit drugs within the last 6 months, frequent medication intake (except oral contraceptives) and confirmed somatic or psychiatric diseases within the last 6 months. Participants had normal or corrected to normal vision. Participants received either course credits toward their research requirements or a monetary incentive of 30 CHF. Prior to the experiment, informed consent conforming to institutional guidelines for human research was obtained from participants. The study was approved by the university's ethics committee and is in accordance with the latest revision of the Declaration of Helsinki (World Medical, 2013).

Stimulus materials and procedure

Using E-Prime (Psychology Software Tools Inc., Pittsburgh, USA) software, three neutral faces taken from the International Affective Picture System (Lang *et al*., 2005; numbers 2372, 2493, 2499) were presented in pseudo-randomized order in eight blocks of 45 repetitions of each face (135 pictures per block, 1080 trials in total, Figure 1) with a stimulus presentation time varying randomly from 700–1300 ms and an interstimulus interval (ISI) varying randomly from 800–1400 ms. Presentation times and ISIs varied randomly to avoid presentation rhythm anticipation. We chose to use three face pictures (including one CS^+ and two CS^-) in order to reduce monotony, to increase discrimination demands and to reduce the strong safe-unsafe polarity that studies using only two CSs have encountered (Lissek *et al*., 2005; Blechert *et al*., 2008). Pseudo-randomization was optimized so that transition frequencies among the three conditions were nearly equal overall (≈320) and within each block (≈45). Transition frequencies were reanalyzed after artifact correction within each of the eight blocks and did not differ between conditions (Ps > 0.424). All faces and pause screens had a 50% grey background and black fixation cross overlaid at the screen center. Participants were asked to keep their gaze fixed on this cross during the presentation blocks. Faces were adjusted to be of the same size, brightness and contrast and filled a large part of the screen (15.3° of 28° horizontal visual angle, 16.6° of 21° vertical visual angle). Slides were presented on a monitor with an 85 Hz refresh rate. Participants were instructed to simply view the pictures. They were informed that sometimes during the experiment they would feel an electric shock, but no other information about the CS–US contingency was given.

During the third, fourth and fifth block (acquisition phase), one of the CS (the 'CS^+') was always followed by a moderate electrical stimulation as US (100% reinforcement) applied to the lower left arm of the subject. This 200 ms current (50 Hz alternating sinusoid ranging from –5 to 5 V) intensity was adjusted individually by each subject within the range from 0.05 to 10 mA at the beginning of the experiment in order to be 'highly unpleasant but not painful'. The US was applied immediately after the offset of the CS+ and did not influence the length of the ISI (one face until the next). The other two faces (the 'CS^{-1}' and the 'CS^{-2}') were followed simply by the ISI and the next trial. During the first and second block (habituation phase), and sixth, seventh and eighth block (extinction phase), no US was presented. The CS–US contingency was balanced over the 24 participants, so that each face was the CS^+ for 8 participants, CS^{-1} for 8 participants and CS^{-2} for the remaining 8 participants. Note that the habituation phase was shorter (two blocks) than both the acquisition and extinction phases (three blocks each).

After each block of trials, a rating sequence was presented during which participants were asked to indicate CS valence ratings and US expectancy of the three CSs on a visual analogue scale presented on the screen. At the end of the acquisition phase, contingency awareness was assessed by presenting the three CSs with the question 'which one of the three pictures had been paired with the electrical stimulation?'.

Apparatus and data preprocessing

The electrical stimulation was delivered using a BIOPAC MP150 amplifier system equipped with a STM100C stimulator module, which was triggered by E-Prime 1.2 (Psychology Software Tools, Pittsburgh, PA, USA) parallel port output. The current was applied between two disposable Electrocardiogram (ECG) electrodes attached to the front side of the lower-left arm with a distance between electrodes of ∼5 cm.

The EEG was collected continuously from the scalp using a 70-channel active electrode system (Biosemi, Amsterdam, Netherlands) without hardware high-pass filtering at a sampling rate of 256 Hz. Absolute electrode offsets were kept below 25 mV, which is appropriate for this type of EEG amplifier. Data preprocessing was performed using the MATLAB (The MathWorks, Inc., Massachusetts, USA)-based software packages EEGLAB (Delorme and Makeig, 2004) and EMEGS (Peyk *et al*., 2011). Continuous data was re-referenced from the Common Mode Sense to Cz and then filtered with a 0.1 Hz high-pass frequency cut-off. To remove eye artifacts independent of stimulus-locked evoked activity, trials for the removal of eye-related artifacts were generated by subdividing all available data (including the period before and after the start of the experiment) into neighboring 1 s epochs and rejecting extremely corrupted epochs automatically in EEGLAB. Default settings for automatic epoch rejection were used [threshold limit microvolt (μV), 1000; probability threshold (s.d.), 5; maximum percentage of trials to reject per iteration, 5]. The remaining epochs (approximately two-thirds of all available epochs) were used to identify eye movement and eye blink components using the default independent component analysis (ICA) algorithm available in EEGLAB.

The three components corresponding to (i) blinks, (ii) horizontal and (iii) vertical eye movements were selected manually for each subject after visual inspection of the component maps and the component activations (Plochl *et al*., 2012). These

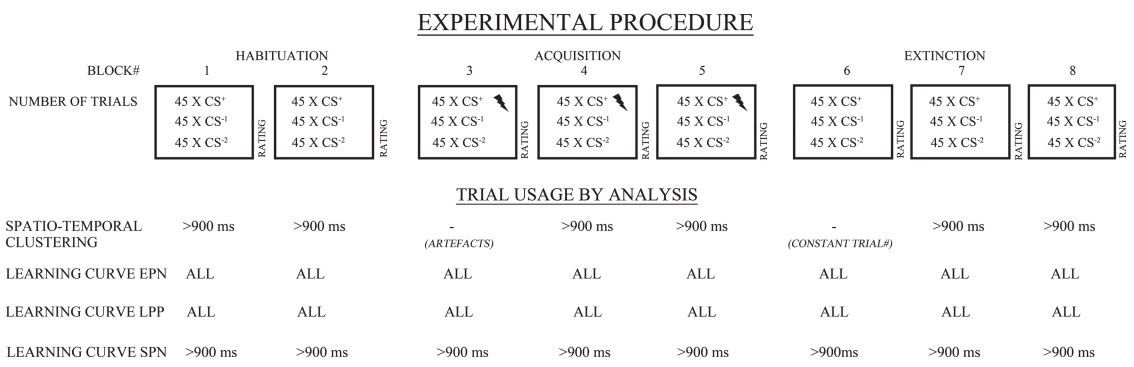

Fig. 1. Schematic representation of the experimental procedure and trial configuration for the performed analyses.

components were removed from the original continuous data and stimulus-locked epochs were extracted as described in the following section. The method for statistical control of artifacts (Junghöfer *et al.*, 2000) was then used for editing and artifact rejection. This procedure (i) detects individual channel artifacts, (ii) detects global artifacts, (iii) replaces artifact-contaminated sensors with spline interpolation statistically weighted on the basis of all remaining sensors and (iv) computes the variance of the signal across trials to document the stability of the averaged waveform. The rejection of artifact-contaminated trials and sensor epochs relies on the calculation of statistical parameters for the absolute measured scalp potential amplitudes over time, their standard deviation over time, the maximum of their gradient over time (first temporal derivative) and the determination of boundaries for each of these three parameters. The number of good trials per average was controlled within every subject by removing randomly selected surplus trials from the cleaner CS-types to match the most contaminated CS-type (this step was performed for both the component identification (see 2.4) and the learning curve reconstruction (see 2.5). Finally, clean data epochs were transformed to an average reference.

ERP component identification and spatio-temporal clustering

In the first step, to identify components sensitive to the experimental manipulations, stimulus-locked epochs were extracted from the EEG data corrected for eye artifacts from 700 ms pre- until 900 ms post-onset of a given stimulus. For this analysis, the trials from all experimental blocks were used, grouped only by CS-type (CS^{+}, CS^{-1} and CS^{-2}) and experimental phase (habituation, acquisition, extinction). To be able to analyze the SPN component, which corresponds to the expectation of the US, from all trials only those with a CS duration of more than 900 ms (~50% of all trials) were selected for this first step of the analysis. The discarded (~50% of the) trials were necessary to induce the expectation of the US around 600 ms after stimulus onset, but were corrupted too heavily by the electrical stimulus and related movement artifacts. At the beginning of the acquisition phase, the EEG of nearly all subjects was confounded by movement artifacts due to the US presentation. To remove these from the event-related averaging and also to keep the number of trials of each experimental phase constant, trials from blocks three and six were excluded for this analysis. This resulted in a maximum of 90 trials per average (Figure 1).

On the averaged waveforms of the remaining trials, repeated measures analyses of variance (ANOVAs), including the factors Block (habituation, acquisition and extinction) and CS-type (CS^{+}, CS^{-1} and CS^{-2}), were calculated for all sensors and time points after picture onset. To correct for multiple comparisons and consider potential deviations from normal distribution, non-parametric spatio-temporal cluster level statistics as suggested by Maris and Oostenveld (2007) were calculated on the *F*-values of the CS-type × Block interaction. Analyses were performed using the MATLAB (Release 2012a)-based software EMEGS (Peyk *et al.*, 2011; www.emegs.org) on 1000 random permutations of the original participant data set. Cluster analyses were conducted using a first-level (sample level) error probability of 0.001. Effects were considered meaningful when emerging in a spatio-temporal cluster achieving a cluster level error probability below 0.001. This very conservative threshold was selected to ensure sufficient effect sizes for the following learning curve reconstruction.

Clusters were identified over the entire response window (0–900 ms after stimulus onset) and cluster masses (the sum of the *F*-values of all samples comprised in a cluster) are reported as test statistic of the non-parametric test. Significant clusters were explored by Sheffé *post hoc* contrasts. Partial η^2 and Cohen's *d* (using an average variance) were calculated as effect size measures where appropriate.

Learning curve reconstruction

In the second step, EEG epochs within each block were divided into subsets of consecutive trials to analyze the temporal dynamics of the learning process. As the analysis of the SPN without the interference of US-induced artifacts was possible only for trials with CS durations >900 ms, the analysis of the EPN and LPP components was separated from the analysis of the SPN components, in order to maximize the number of available trials for the EPN and LPP components (the corresponding shorter stimulus-locked epochs comprised 200 ms before and 600 ms after stimulus onset).

For the EPN and LPP data, starting with an average across each of the eight experimental blocks (Figure 1), epochs of each CS-type within each of the eight blocks were divided into an increasing number of subsets of consecutive trials, starting with 2 subsets and going to 4, 8, until a maximum of 16 subsets (see the 'trial divider' in Figures 4–6). The number of trials per subset was thus decreased from 45 to (rounded to the closest integer) 23, 11, 6 and 3 trials per subset (see the 'number of trials' in Figures 4–6). As the number of 45 trials was not evenly dividable by 2, subset limits were calculated as floating point numbers, and trials per average could vary by one trial across subsets. These trial numbers represented the maximum number of theoretically available trials. A small fraction of trials

was contaminated by artifacts, as can be seen for instance by the missing data in Figures 4–6 at the start of the acquisition phase. However, the overall data quality was very high, allowing for the analysis of averages of only three trials. The theoretically possible analysis based on single trials was performed but will not be reported due to more than 50% missing data. The habituation ERP of the overall component analysis served as a reference ERP, and the difference to this waveform was calculated for every average and CS-type. CS^+-specific processing was tested by calculating average regional amplitudes of the different ERPs corresponding to the spatio-temporal clusters found in the overall component identification step. These regional amplitudes were compared using the maximum (least significant) *P*-value of two dependent sample *t*-tests, comparing the CS^+ with each of the CS^- (Figures 4–6). The minimum of the absolute (non-directional) value of two Cohen's *ds* (using an average variance) corresponding to these two *t*-tests was used as effect size measure.

This combination of two *t*-tests was preferred over a custom *F*-contrast (with for instance two negative weights of −1 and one positive weight of +2) since it requires more strictly that the CS^+ differs from both CS^- in the expected direction (more negative for EPN and SPN and more positive for the LPP, note that the *F*-contrast can become significant with only one comparison showing the expected pattern). *T*-tests were performed without correction for multiple comparisons to maximize the sensitivity of the analyses. As a result, findings might involve false positives and should be interpreted with caution.

As the analysis of the SPN component could be performed with only about half the trials of the EPN/LPP analysis (23 trials per block), the number of trials within each block could only be divided into 2, 4 and 8 subsets of consecutive trials per block (see the 'trial divider' in Figure 6; corresponding to 11, 6 and 3 trials per subset; see the 'number of trials' in Figure 6). The highest temporal resolution of 16 subsets could not be achieved with the SPN data. CS^+-specific processing was tested in the same way as described for the EPN/LPP data.

Results of different temporal resolutions are reported using the 'trial divider' in Figures 4–6 as subscripts (e.g. P_1 indicating a *P*-value from the lowest temporal resolution, P_{16} indicating a *P*-value from the highest temporal resolution).

To identify differences in the learning curves between components, in the first step the (negative) EPN and SPN regional amplitudes were multiplied by −1 to obtain uniform (positive) polarities. To compensate for the different effect sizes, scaling values were obtained by calculating regional amplitude averages across all subjects and selecting the maximum absolute value of these grand averages during the acquisition in the lowest temporal resolution (trialdivider 1) separately for the EPN, LPP and SPN component. These scaling values always corresponded to the CS^+-condition since this condition provided the largest absolute amplitudes, but from different trial sets across components (trial set 4 for the EPN, 5 for the LPP and 4 for the SPN; see Figures 4, 5 and 6), illustrating slightly different temporal dynamics. The regional amplitudes of the three components of all subjects of all temporal resolutions were then divided by the corresponding scaling values in order to scale each component to a maximum absolute grand average amplitude of +1 at the lowest temporal resolution. All temporal resolutions were scaled by the same values from the lowest temporal resolution to make the scaling procedure independent of the (decreasing) signal-to-noise ratio. In the second step, subsets of trials of each component showing CS+-specific processing (Figures 4–6) were compared with the other components using these transformed regional amplitudes. To that effect, the CS^--CS^+ difference of the lesser of the two *t*-tests (the one with the larger *P*-value) within one component (described above for detecting CS^+-specific processing) was subtracted from the CS^+-CS^--difference from the lesser of the two *t*-tests (the one with the larger *P*-value) from the other component for each subject. The resulting values were then *t*-tested against zero. This is equivalent to a custom *F*-contrast with the first component weights of 1 and −1, and the second component weights of −1 and 1 for CS+ and CS−, respectively, using the weaker of the two CS^+-CS^- contrasts (the one with the larger *P*-value) within each component. Namely, for every trial set where CS^+-specific processing was found, we calculated three *t*-tests: 1.) EPN(CS^+-$CS^{-(weaker)}$) minus LPP(CS^+-$CS^{-(weaker)}$) (figure 4), 2.) EPN(CS^+-$CS^{-(weaker)}$) minus SPN(CS^+-$CS^{-(weaker)}$) (also figure 4) and 3.) LPP(CS^+-$CS^{-(weaker)}$ minus LPP(CS^+-$CS^{-(weaker)}$ (figure 5). These tests were performed in a two-tailed fashion, but only positive significant results were found (EPN>LPP, EPN>SPN, and LPP>SPN). Partial eta-squared values are reported as effect sizes.

Analysis of subjective ratings

US expectancy ratings and CS valence ratings were analyzed in a manner matching the EEG data: first, a standard ANOVA analysis, using an average of only the two final blocks of each experimental phase to get a picture of the status at the end of each experimental phase; second, a per-time-point combination of two *t*-tests to explore the changes during the blocks in more detail. The standard ANOVA analysis included the factors Block (habituation, acquisition and extinction) and CS-type (CS^+, CS^{-1} and CS^{-2}). Significant effects were explored using Sheffé *post hoc* contrasts and partial η^2 and Cohen's *d* were calculated as effect size measures where appropriate. The per-time-point analysis was based on the maximum (least significant) *P*-value of two dependent sample *t*-tests, as described above (Figure 7). As for the EEG data, the minimum of the two Cohen's *ds* (using an average variance) corresponding to these two *t*-tests was used as effect size measure.

Results

Component identification and spatio-temporal clustering

The cluster level analysis revealed five significant clusters of the Block × CS-type interaction (in size descending order): a right-central cluster from 535–900 ms (cluster mass 11 082.9, $P < 0.001$), a frontal cluster from 727–900 ms (cluster mass 5402.3, $P < 0.001$), a parietal cluster from 340–594 ms (cluster mass 2819.5, $P < 0.001$), an occipito-temporal cluster from 195–324 ms (cluster mass 1419.2, $P < 0.001$) and a central cluster from 230–309 ms after stimulus onset (cluster mass 538.6, $P < 0.001$, Figure 2). The temporal and spatial characteristics of the clusters 1, 3 and 4 strongly resembled previously reported components in motivated attention research, as did the corresponding ERP waveforms (Figure 3). Therefore, these will be referred to as SPN, LPP and EPN clusters in the following. Clusters 2 and 5 were temporally very similar to clusters 1 and 3, but with lower clusters masses. As the transformation to an average reference requires the integral of all head surface potentials to be zero at every time point, such mirror effects are very commonly encountered in ERP studies applying an average reference (e.g. Schupp *et al.*,

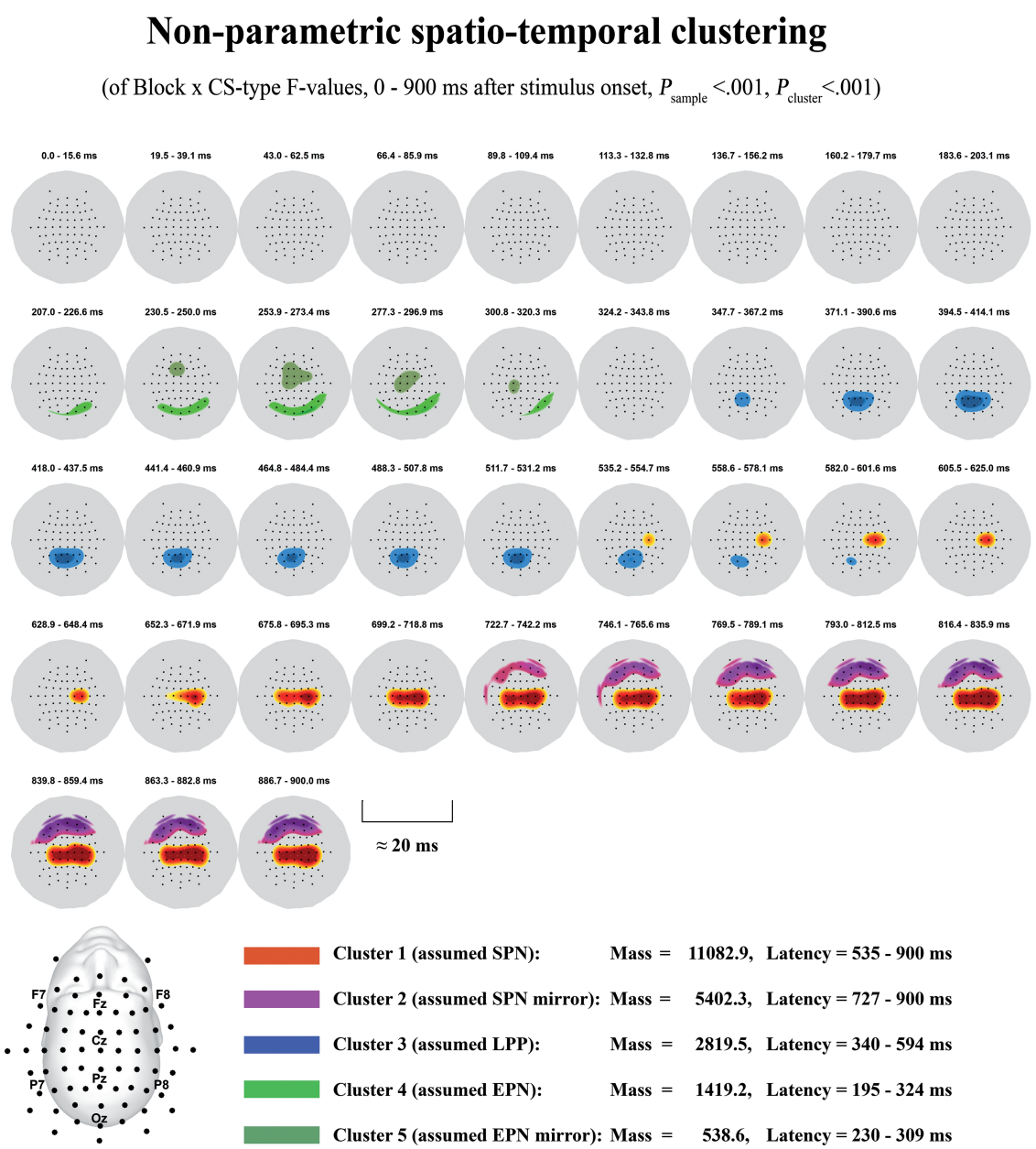

Fig. 2. Planar projected topographical maps of the spatio-temporal clustering. Dots indicate sensor positions as seen from above with anterior sensors at the top of the page and left sensors on the left side. Colored areas indicate samples within a significant spatio-temporal cluster. Each map corresponds roughly to 20 ms, each row to a little more than 200 ms.

2004; Peyk *et al.*, 2009). Consequently, these two clusters were not investigated any further.

Post hoc comparisons exploring the sample-level effect of the EPN cluster [$F(4,92) = 22.9$, $P < 0.001$, $\eta_p^2 = 0.492$] revealed that during acquisition the CS^+ evoked a lower EPN amplitude than the CS^{-1} ($P < 0.001$, $d = -0.821$) and CS^{-2} ($P < 0.001$, $d = -1.039$). The EPN amplitude elicited by the CS^+ was also lower during acquisition by comparison with the habituation phase ($P < 0.001$, $d = -0.826$) and returned to its initial amplitude during extinction [$CS^+_{HAB} \neq CS^+_{EXT}$: $P = 0.232$, non-significant(NS)]. CS-types did not differ significantly in the habituation ($Ps > 0.995$) and extinction ($Ps > 0.094$) phase nor did the CS^{-1} and CS^{-2} differ across phases ($Ps > 0.458$).

Similarly, *post hoc* comparisons exploring the sample level effect of the LPP cluster [$F(4,92) = 17.4$, $P < 0.001$, $\eta_p^2 = 0.433$] showed that during the acquisition phase the CS^+ evoked a greater LPP component than the CS^{-1} ($P < 0.001$, $d = 1.490$) and CS^{-2} ($P < 0.001$, $d = 1.149$). The LPP component elicited by the CS^+ in the acquisition was greater than in the habituation ($P < 0.0001$, $d = 1.078$). CS-types did not differ significantly on the LPP component in the habituation ($Ps > 0.996$) and extinction ($Ps > 0.995$) phase nor did the CS^{-1} and CS^{-2} differ across phases ($Ps > 0.868$). For the LPP component just as for EPN, the CS^+ returned to its initial amplitude during extinction ($CS^+_{HAB} \neq CS^+_{EXT}$: $P = 1.0$, *NS*).

Finally, sample level *post hoc* tests of the SPN cluster [$F(4,92) = 32.96$, $P < 0.001$, $\eta_p^2 = 0.595$] indicated that only the CS^+ during acquisition differed from all other cells (all $Ps < 0.001$, all $ds < -1.490$), showing a negative slow potential. No other comparisons were significant (all $Ps > 0.05$).

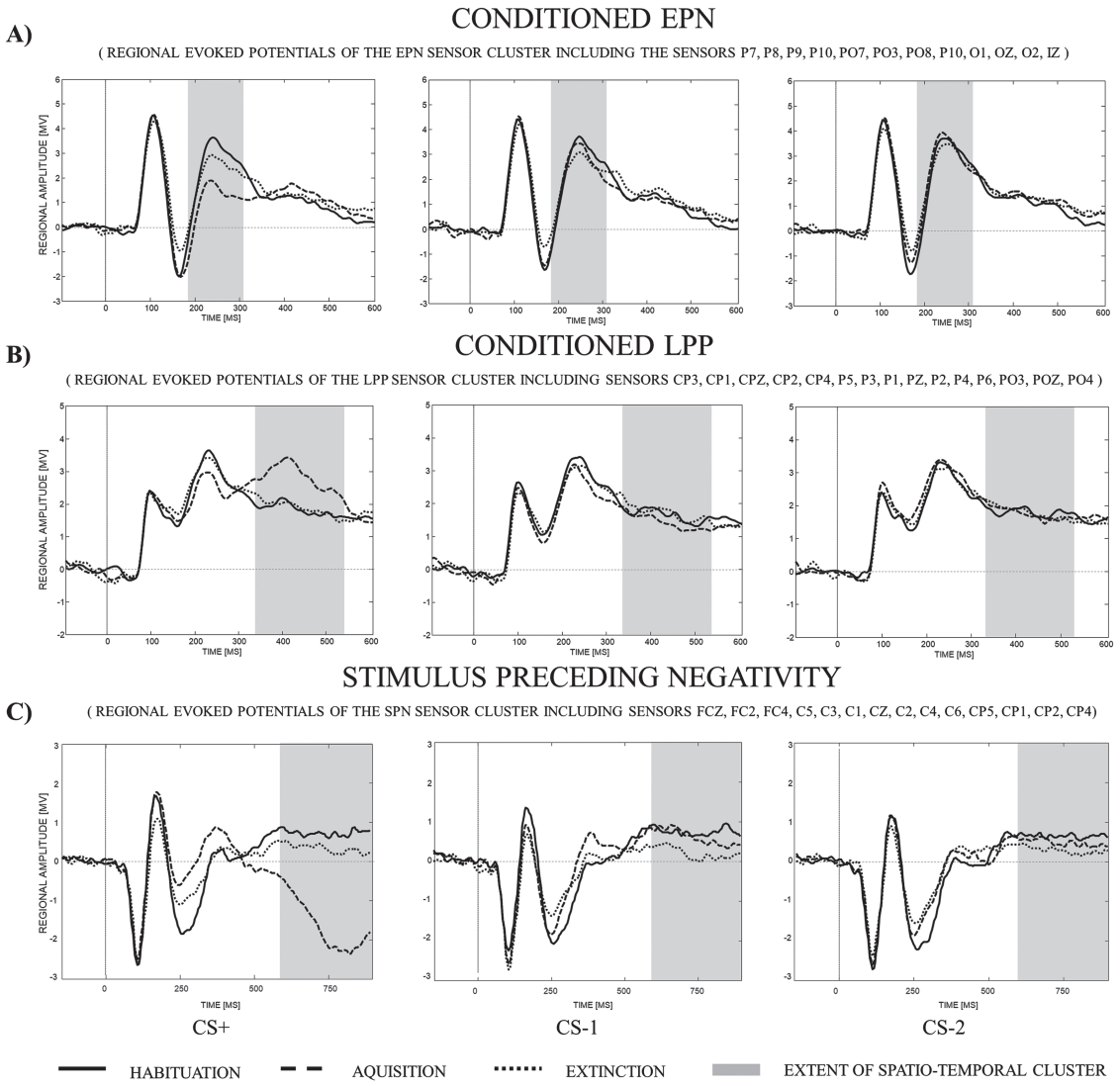

Fig. 3. Regional ERP grand average waveforms elicited by the different stimuli: CS^{+}, CS^{-1} and CS^{-2}. Each line represents a different phase of the fear conditioning paradigm (habituation, acquisition and extinction).

Learning curve reconstruction

EPN cluster. The learning curve reconstruction of the EPN effect demonstrated a CS^{+}-specific processing change triggered during the acquisition phase at all temporal resolutions, even at the highest with a maximum number of only three good trials per average (subscripts denote the trial divider: minimal $P_1 < 0.01$, minimal $P_2 < 0.001$, minimal $P_4 < 0.01$, minimal $P_8 < 0.01$, minimal $P_{16} < 0.001$, maximal absolute $d_1 = 0.886$, maximal absolute $d_2 = 0.990$, maximal absolute $d_4 = 0.743$, maximal absolute $d_8 = 0.946$, maximal absolute $d_{16} = 0.872$). The percentage of trial subsets in the acquisition phase exhibiting this pattern decreased as the number of trials per subset was reduced, from initially 100% at 45 trials per subset to 83, 64, 57 and 20% at 23, 11, 6 and 3 trials per subset, respectively. However, even at the highest temporal resolution it lay far above chance level.

This processing change was reversed during the extinction phase (Figure 4). The acquisition took around 30 trials to fully develop ($N_1 = 0$, $N_2 = 23$, $N_4 = 33$, $N_8 = 36$, $N_{16} = 32$ with N indicating the trial position of the first significant subset in the acquisition), and the extinction process took just as many trials if not a little more ($N_1 = 45$, $N_2 = 45$, $N_4 = 0$, $N_8 = 36$, $N_{16} = 33$ with N indicating the trial position of the last significant subset).

LPP cluster. Just as for the EPN cluster, the learning curves of the LPP effect showed CS-type-dependent modulation at all temporal resolutions (minimal $P_1 < 0.0001$, minimal $P_2 < 0.0001$, minimal $P_4 < 0.0001$, minimal $P_8 < 0.001$, minimal $P_{16} < 0.001$). However, it was considerably larger than the EPN effect as indicated by the larger absolute effect sizes (maximal $d_1 = 1.327$, maximal $d_2 = 1.209$, maximal $d_4 = 1.104$, maximal $d_8 = 1.109$, maximal $d_{16} = 1.268$) and more sustained as indicated by higher percentages of subsets in the acquisition phase exhibiting this pattern: a 100% at 45 and 23 trials per subset, 91% at 11, 60% at 6 and still 31% at 3 trials per subset.

Significant modulation of the LPP during the acquisition could be seen already after about 15 trials ($N_1 = 0$, $N_2 = 0$, $N_4 = 11$, $N_8 = 18$, $N_{16} = 15$). Assuming the one late outlier at the highest temporal resolution was due to the low signal to noise ratio rather than a true effect, this quick acquisition was also accompanied by a quick extinction, with CS+-specific processing only visible within the first few trials in the extinction phase ($N_1 = 45$, $N_2 = 23$, $N_4 = 6$, $N_8 = 6$, $N_{16} = 0$; Figure 5).

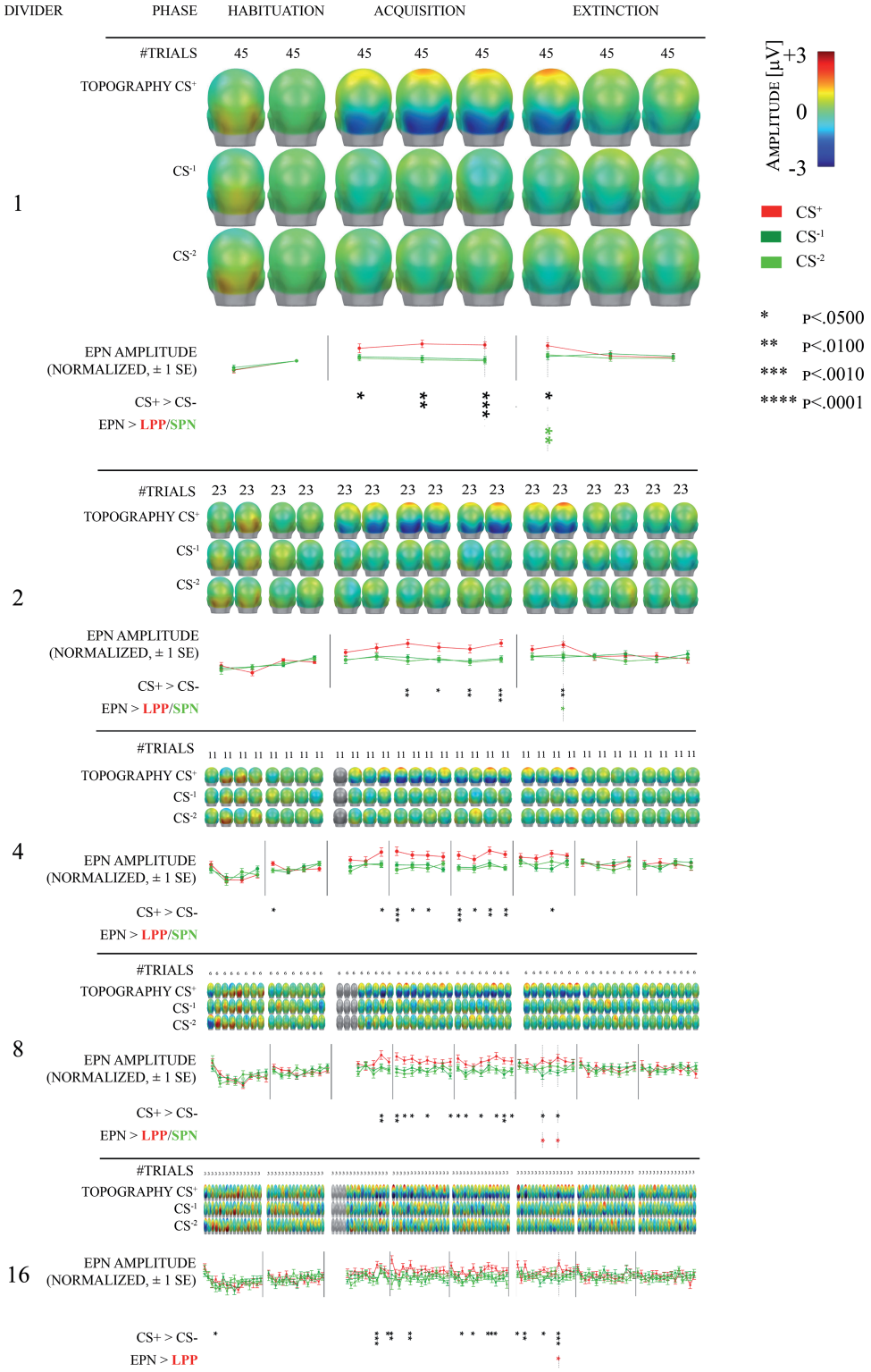

Fig. 4. Learning curve reconstruction of the EPN component by consecutive subdivision of the trials per average. The three rows of topographical maps (back view on realistic head surface model) per subdivision depict the temporal trajectory of the three conditioned stimuli (CS^{+}, CS^{-1} and CS^{-2}) relative to the end of the habituation phase. Line plots indicate regional amplitudes corresponding to the spatio-temporal extent of the EPN cluster (scaled and normalized to be comparable across components). Black stars indicate the level of significance of the *t*-tests for CS+-specific processing within the EPN component, colored stars indicate the level of significance of the *t*-tests for the between component comparisons. Error bars correspond to ±1 standard error *(SE)*.

SPN cluster. Although temporal resolution was more limited, the learning curve reconstruction in the SPN window showed CS-type-dependent modulation even more clearly than for the LPP component (minimal $P_1 < 0.0001$, minimal $P_2 < 0.0001$, mini-

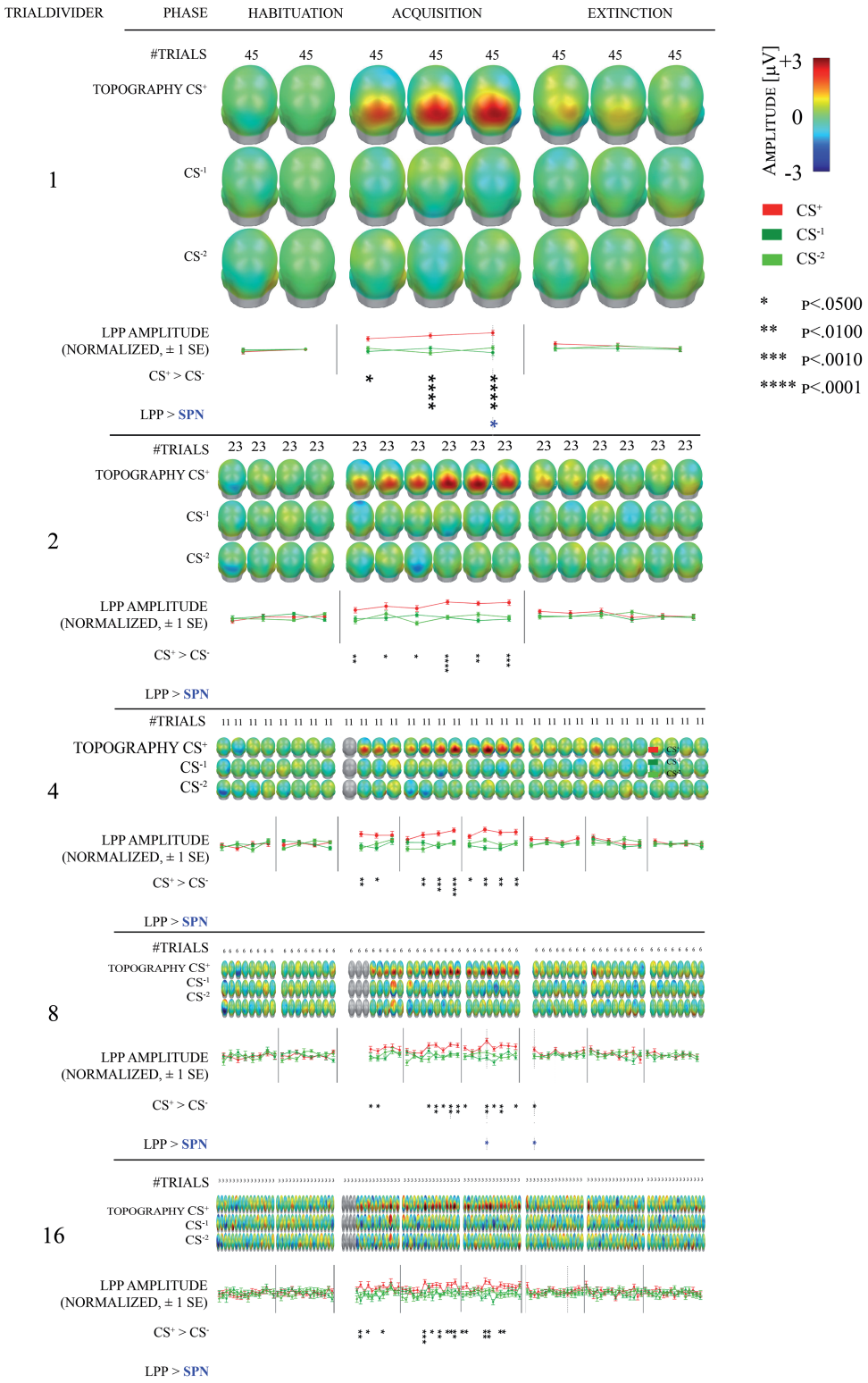

Fig. 5. Learning curve reconstruction of the LPP component by consecutive subdivision of the trials per average. The three rows of topographical maps (back view on realistic head surface model) per subdivision depict the temporal trajectory of the three conditioned stimuli (CS^{+}, CS^{-1} and CS^{-2}) relative to the end of the habituation phase. Line plots indicate regional amplitudes corresponding to the spatio-temporal extent of the LPP cluster (scaled and normalized to be comparable across components). Black stars indicate the level of significance of the *t*-tests for CS+-specific processing within the LPP component, colored stars indicate the level of significance of the *t*-tests for the between component comparisons. Error bars correspond to ±1 *SE*.

mal $P_4 < 0.0001$, minimal $P_8 < 0.001$) with even larger absolute effect sizes (maximal absolute $d_1 = 1.434$, maximal absolute $d_2 = 1.526$, maximal absolute $d_4 = 1.393$, maximal absolute $d_8 = 0.966$). Significant SPN modulation could be seen very early

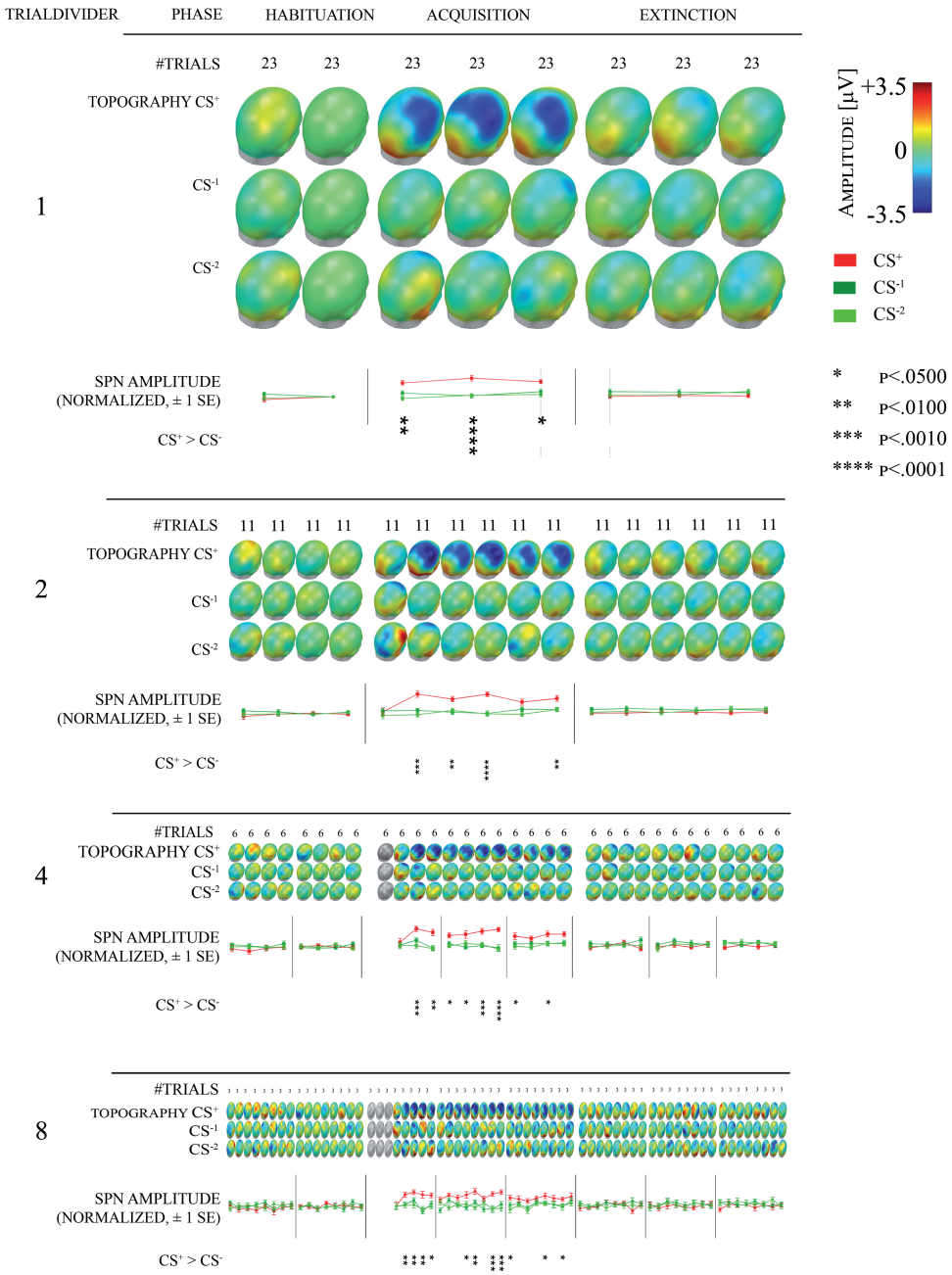

Fig. 6. Learning curve reconstruction of the SPN component by consecutive subdivision of the trials per average. The three rows of topographical maps per subdivision (back right view on realistic head surface model) depict the temporal trajectory of the three conditioned stimuli (CS^+, CS^{-1} and CS^{-2}) relative to the end of the habituation phase. Line plots indicate regional amplitudes corresponding to the spatio-temporal extent of the SPN cluster (scaled and normalized to be comparable across components). Stars indicate the level of significance of the *t*-tests for CS+-specific processing within the SPN component. Error bars correspond to ±1 *SE*.

in the acquisition ($N_1 = 0$, $N_2 = 11$, $N_4 = 12$, $N_8 = 12$). Interestingly, the SPN effect seems to have faded before the end of the acquisition (Figure 6), and no averages revealed CS^+-specific processing anymore in the extinction ($N_1 = 0$, $N_2 = 0$, $N_4 = 0$, $N_8 = 0$).

Component comparisons

Cross-component contrasts revealed some significant differences between components at the end of the acquisition and the beginning of the extinction (subscripts denote the trial divider and corresponding trial set number). The CS+-specific EPN component was significantly larger than the CS+-specific SPN component at the beginning of the extinction phase: $T_{1,Set6}(23) = 2.843$, $P < 0.01$, $\eta^2_p = 0.26$; $T_{2,Set12}(23) = 2.651$, $P < 0.05$, $\eta^2_p = 0.234$; Figure 4). It was also larger than the CS+-specific LPP component at some trial sets at the beginning of the extinction at higher temporal resolutions [$T_{8,Set44}(23) = 2.188$, $P < 0.05$, $\eta^2_p = 0.172$; $T_{8,Set46}(23) = 2.519$, $P < 0.05$, $\eta^2_p = 0.216$; $T_{16,Set92}(23) = 4.368$, $P < 0.001$, $\eta^2_p = 0.453$]; however, this effect failed to reach significance at lower temporal resolutions [$T_{1,Set6}(23) = 1.494$, $P = 0.148$, $\eta^2_p = 0.089$; $T_{2,Set12}(23) = 1.716$, $P = 0.1$, $\eta^2_p = 0.113$]. The CS+-specific LPP component was significantly larger than the CS+-specific SPN

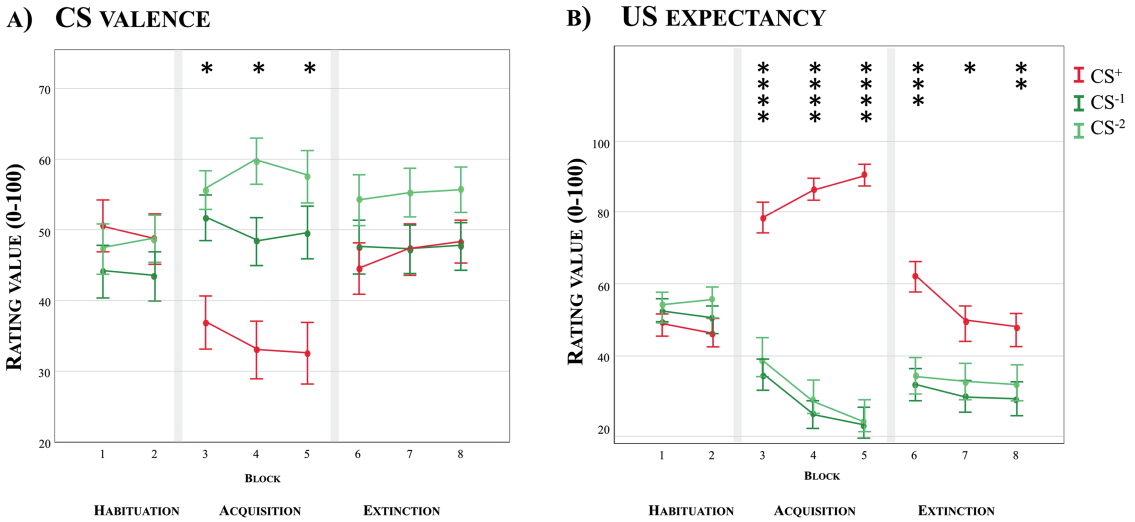

Fig. 7. Means and standard errors of online valence and US expectancy ratings for the CS^{+}, CS^{-1} and CS^{-2} during the eight blocks of the experiment. Low ratings on the valence scale represent more negative stimulus valence. High ratings on the US expectancy scale represent a higher expectancy of a shock US following the CS. Stars indicate the level of significance of the *t*-tests for CS+-specific rating (* <0.05, ** <0.01, *** <0.001, **** <0.0001).

component towards the end of the acquisition [$T_{1,Set5}(23) = 2.144$, $P < 0.05$, $\eta^2_p = 0.167$; $T_{8,Set36}(23) = 2.248$, $P < 0.05$, $\eta^2_p = 0.18$] and beginning of extinction [$T_{8,Set41}(23) = 2.541$, $P < 0.05$, $\eta^2_p = 0.219$].

CS valence

CS valence ratings (Figure 7a) in the overall ANOVA followed the pattern of the EEG data, a significant Block × CS-type interaction [(Block × CS-type $F(4,92) = 15.20$, $P < 0.0001$, $\eta_p{}^2 = 0.398$; main effects discarded: Block $F(2,46) = 2.55$, *NS*, $\eta_p{}^2 = 0.100$; CS-type $F(2,46) = 2.76$, *NS*, $\eta_p{}^2 = 0.108$], with lower valence ratings for the CS^{+} during the acquisition phase only ($CS^{+}_{Acq} < CS^{-1}_{Acq}$: $P < 0.0001$, $d = -0.863$; $CS^{+}_{Acq} < CS^{-2}_{Acq}$: $P < 0.0001$, $d = -1.268$).

The temporal analysis of this rating response confirmed that it was quickly learnt, being apparent already in the first block of the acquisition and increasing only slightly in the following acquisition blocks. Its extinction was similarly fast being almost completed already at the end of the first extinction block.

US expectancy

US expectancy ratings (Figure 7b) in the overall ANOVA also revealed a strong Block × CS-type interaction [Block × CS-type $F(4, 92) = 73.61$, $P < 0.0001$, $\eta_p2 = 0.762$; main effects discarded: Block $F(2,46) = 6.29$, $P_{GG} < 0.05$, $\eta_p2 = 0.215$; CS-type $F(2,46) = 26.32$, $P < 0.0001$, $\eta_p2 = 0.534$], with greater expectancy of the US after CS+ during the acquisition phase (CS^{+}Acq > CS^{-1}Acq: $P < 0.0001$, $d = 3.705$; CS^{+}Acq > CS^{-2} Acq: $P < 0.0001$, $d = 3.822$), but also a medium expectancy during the extinction (CS^{+}Acq ≠ CS^{+}Ext: $P = 0.636$, *NS*; CS^{+}Ext > CS^{-1}Ext: $P < 0.01$, $d = 0.839$; CS^{+}Ext > CS-2Acq: $P < 0.05$, $d = 0.639$). The two CS^{-} were similar in all experimental phases, decreasing from a medium expectancy level during the habituation to a significantly lower level during acquisition and extinction (CS^{-1}Acq < CS^{-1}Hab: $P < 0.0001$, $d = {}^{-}1.492$; CS^{-1}Acq ≠ CS^{-1}Ext: $P < 1.0$, *NS*; C^{S-2}Acq < CS^{-2}Hab: $P < 0.0001$, $d = {}^{-}1.588$; CS^{-2}Acq ≠ CS^{-2}Ext: $P < 0.995$, *NS*).

Temporal analysis of this rating effect showed that it was apparent already in the first block of the acquisition. However, it continued to increase throughout the acquisition phase. Its extinction was only partial, with the CS^{+} returning to its initial level in the second block of the extinction, but the two CS^{-} continuing to signal relative safety throughout all extinction blocks.

Contingency awareness

All participants were able to correctly identify the CS^{+} as the stimulus that was coupled with the electric shock at the end of the acquisition phase.

Discussion

The present study aimed to investigate the temporal dynamics of ERPs of motivated attention during acquisition and extinction of social cue oriented conditioned fear in healthy participants. ERP data revealed strong and reliable modulations by the fear learning procedure. Without any a priori constraints, the non-parametric cluster analysis identified effect clusters corresponding to the temporal and topographical properties of the EPN, LPP and SPN components.

During the acquisition, the fear CS^{+} quickly started to elicit an enhanced EPN, LPP and SPN, compared to the end of the habituation phase. These changes were reversed by extinction indicating that a fast readjustment of motivation had taken place. The present analysis of temporal dynamics illustrates the speed by which the cortical system adapts to changes of motivational relevance and emphasizes the high plasticity of learning and extinction of social fear in healthy humans.

These results were paralleled by the valence and US expectancy ratings, indicating that during the acquisition phase participants became to dislike the CS^{+} and expect it to be followed by a US. During extinction, valence returned to its initial values and the US expectancy decreased significantly. Self-report measures indicate that the fear conditioning manipulation was successful.

The here reported results support and extend previous findings on ERP activity in fear conditioning paradigms (Flor *et al.*, 2002; Wong *et al.*, 2004; Lee *et al.*, 2009; Panitz *et al.*,

2015). The enhanced EPN suggests that attention was selectively and automatically oriented to the newly learned social threat cue. The enhanced LPP activity during the acquisition phase seems to indicate that motivation towards this social threat cue was changed and that in comparison with the other non-threat social cue, it had a stronger representation in working memory, crucial for decision making and adaptive behavior. The enhanced anticipatory attention (SPN) can be assumed to have helped increase the cognitive processing efficiency and prepare the body to swiftly react to the expected noxious stimulus.

Additionally and more interestingly, some evidence for specific temporal dynamics of the three investigated components were found.

The extinction process was longer for the EPN than for the LPP and SPN, and the SPN seems to have slightly faded already before the beginning of the extinction. These distinct temporal patterns may give hints as to the functional meaning of the corresponding ERP components and their role in fear learning. The exact neural mechanism of altered processing of affective visual stimuli is still unknown. The traditional concept of reentrant projections from the amygdala into secondary visual cortices, which had been adopted from findings in the auditory system by LeDoux (2000), can be considered as a model of automatic bottom-up processes. However, there is also convincing evidence for top-down modulation of the processing of emotional stimuli by prefrontal cortex areas (Bar *et al*., 2006; Steinberg *et al*., 2012), and it seems likely that multiple processing routes are involved in the visual perception of emotional stimuli (Schirmer and Adolphs, 2017).

Complicating things further, a large animal and human literature supports the idea that fear acquisition memory and fear extinction memory are separate learning processes rather than the formation and undoing of the same single memory (Bouton, 2002), with acquisition involving the creation of associations in the lateral amygdala and extinction consisting in the creation of amygdala-inhibiting projections from (medial) prefrontal areas (Pape and Pare, 2010). Moreover, it is still unknown whether early ERP effects are realized through plasticity in these brain areas or through plasticity of the sensory cortex itself (Weinberger, 2004; Stolarova *et al*., 2006; Keil *et al*., 2007).

Nonetheless, the present data raises the question whether the EPN stage of neurocognitive processing adapts with greater inertia during fear learning than the LPP or SPN stage. Given that all subjects realized the CS–US contingency easily, it can be hypothesized that a cognitive high-level learning more closely related to the LPP and SPN components, may have been quicker than an intuitive and more automated learning more closely related to the EPN. If this was true, it could also be concluded that top-down control had only a limited effect on early processing stages. This interpretation, however, needs to be made with caution, since during the acquisition no significant contrasts between components were found to substantiate this premise.

During extinction, the differences between components were more consistent. The EPN effect was sustained for a part of the early extinction. In light of the results by Schupp *et al*. (2006b), who have shown that the differential processing of a priori emotional *vs* neutral stimuli, as reflected by the EPN, does not show signs of habituation, the present results could have been expected. However, the fact that the EPN in our study eventually does disappear seems to suggests that the stimuli used by Schupp and colleagues—highly arousing emotional pictures including mutilations and explicit erotic scenes—may be overlearned or biologically predisposed to such an extent that they do not cease to draw attention. Assuming that prefrontal evaluative areas are located upstream of secondary visual areas in processing the CS, it can be hypothesized that the disappearance of the EPN during extinction was the result of the formation of new inhibitory amygdala connections that counteracted the previous fear associations. In this case, the EPN disappearance could be considered a marker for the strength of this new inhibitory memory. If, however, the detection of the motivationally relevant CS was based on plasticity of the visual cortex itself, the fading EPN would be an index of this change.

The extinction of the LPP effect seemed to be somewhat faster than for the EPN effect, and here again, it could be hypothesized that a cognitive high-level process may have been ahead of an intuitive and more automated response. The CS valence ratings seem to support this view, showing a decrease for the CS^+ only during the acquisition (Figure 7).

Finally, the SPN component faded before the end of the acquisition. This might be explained by a consolidated learning of the contingency: a recent study with a feedback task has shown that the SPN component is not only influenced by relevance but also by uncertainty, whereby when uncertainty of the outcome was low, no anticipatory SPN modulation could be seen (Walentowska *et al*., 2017). The same effect might apply to the present data, where uncertainty of the outcome (electroshock) following a CS+ would decrease throughout the acquisition. However, response fading during acquisition has been found with other measures as well. In a study of early evoked oscillatory response in a classical conditioning fear acquisition, Keil *et al*. (2007) found that while there was an increase of the CS+ evoked oscillatory response and dislike ratings, a decrease in the startle response was present. A similar pattern was found by Bacigalupo and Luck (2018), who have shown that while conditioning effects were visible in LPP and self-report during the entire acquisition, electrodermal activity (EDA) responses would greatly decrease after the first block. These effects on physiological measures like eye blink startle and EDA, which are known to habituate strongly, illustrate that in standard classical conditioning paradigms, effects of US uncertainty and habituation might be confounded.

It also seems worth mentioning, that CS^+-specific ERP components were not steadily observed once they were first activated during the acquisition. They were repeatedly turned on and off for a number of consecutive trials. This could be a mere consequence of the low signal-to-noise ratio, but if these shifts were replicated in future studies, they could turn out to be indicators of functional entities in the learning process.

Despite the fact that in the present study we did not find any CS-type-dependent processing effects at earlier latencies than the EPN, our results do not contradict studies that do find these effects (Stolarova *et al*., 2006; Steinberg *et al*., 2012; Mueller *et al*., 2014). Since we selected ERP components based on the two final blocks in each phase, by default we could not find transient effects only visible in early acquisition and extinction. One could even argue that such transient effects should reflect the learning process more directly, whereas the sustained later components that we investigated can be considered to be a 'product' of the learning process. It is also important to bear in mind that due to our analysis approach based on fixed spatio-temporal clusters, we may have missed processing changes that involve changes in latency and topographical properties of the investigated components. Moreover, we used highly conservative alpha error thresholds in the cluster analysis. This was necessary to select large effects that could be temporally explored. However, using these settings we may have overlooked smaller but widespread effects (Maris and Oostenveld, 2007).

The present study makes an important contribution to research in the field of motivated attention in fear learning and extinction, as it quantifies the temporal unfolding of three established ERP components elicited by social stimuli during a classical conditioning paradigm. The here reported effects have a large size (EPN, η_p^2 = 0.492; LPP, η_p^2 = 0.433; SPN, η_p^2 = 0.595) and can reliably reflect changes in motivational relevance even at a lower number of trials per average. Results may foster development of novel ERP fear-conditioning paradigms with low numbers of trials per phase or assessment of the learning trajectory for different neurocognitive processes with high temporal resolution. This may allow important advances in a highly productive research field where typical objective markers of conditioning/extinction are peripheral psychophysiology measures like EDA that do not adequately map into the crucial neurocognitive domain during fear learning (Lonsdorf *et al*., 2017).

Since disturbed fear-learning processes might be responsible for the development and/or maintenance of mental disorders, it may be especially interesting to conduct similar studies with specific groups of patients. For instance, patients with anxiety disorders show enhanced resistance to fear extinction (Blechert *et al*., 2007; Michael *et al*., 2007; Wessa and Flor, 2007; Duits *et al*., 2015) and an impaired fear extinction has been shown to predict the onset of posttraumatic stress disorder (Lommen *et al*., 2013; Zuj *et al*., 2016). Moreover, an attentional bias towards threat stimuli can be found in patients with anxiety disorders (Michalowski *et al*., 2009; Yoon *et al*., 2016). Further exploration of anomalies in fear-conditioning processes using ERP components of motivated attention might contribute to better understanding the mechanisms of learning and extinction of fear memories in anxiety and other mental disorders. Additionally, monitoring EPN, LPP and SPN activity elicited by relevant disorder-related stimuli could serve as a measure of treatment efficiency.

To prevent some of the limitations of EEG analysis in fear conditioning studies, the current study used an optimized design, which is still very similar to the commonly used differential conditioning paradigms (Lonsdorf *et al*., 2017). We used a larger number of trials with short CS time presentation and employed a passive viewing design to avoid confusion of sensory CS processing with task-related activations. In order to focus on the EEG signal, additional physiological measures were not acquired (e.g. EDA, fear-potentiated startle and heart rate). Future studies may extend the current findings to a purely social paradigm, e.g. by using videos of aversive social encounters instead of electrical shock as US (Wiggert *et al*., 2016). Social 'injury' (e.g. blame and rejection) may be processed differently from physical injury represented by electric shock. Furthermore, to better understand the specificities of social-cognitive processing, it would be recommendable to compare fear conditioning to social and non-social stimuli.

It has been shown that although extinction of fear is usually easy to be momentarily learned, return of fear is likely to occur, indicating difficulties with persistent long-term fear extinction (for a review, see Vervliet *et al*., 2013). It is a limitation of the present study that it only focused on fear extinction and it is recommended that future studies include a return of fear test in their experiment. Assuming that late ERP components are more directly linked to contingency awareness than earlier ones, future studies may also manipulate contingency awareness systematically to possibly find some answers on the long debate on contingency awareness in fear conditioning (e.g. Lovibond *et al*., 2011).

The current study shows that learning and extinction of social cue-oriented fear in healthy humans has strong neurocognitive correlates and is a fast and highly plastic process. ERPs of motivated attention appear to be particularly stable markers of these processes and are visible even at higher temporal resolutions, allowing development of novel and advanced paradigms in the study of fear conditioning. The present results can serve as an orientation in the optimization of such study designs. Further studies are necessary in order to better clarify the neuronal substrates of the temporal dynamics of the ERPs of motivated attention in social fear conditioning observed here.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgements

We gratefully acknowledge the assistance of Markus Erni during data collection.

Conflict of interest. None declared.

References

Baas, J.M., Kenemans, J.L., Bocker, K.B., Verbaten, M.N. (2002). Threat-induced cortical processing and startle potentiation. *Neuroreport*, **13**(1), 133–7.

Bacigalupo, F., Luck, S.J. (2018). Event-related potential components as measures of aversive conditioning in humans. *Psychophysiology*, **55**(4):e13015.

Bar, M., Kassam, K.S., Ghuman, A.S., *et al*. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(2), 449–54.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., Muller, K.R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage*, **56**(2), 814–25.

Blechert, J., Michael, T., Vriends, N., Margraf, J., Wilhelm, F.H. (2007). Fear conditioning in posttraumatic stress disorder: evidence for delayed extinction of autonomic, experiential, and behavioural responses. *Behaviour Research and Therapy*, **45**(9), 2019–33.

Blechert, J., Michael, T., Williams, S.L., Purkis, H.M., Wilhelm, F.H. (2008). When two paradigms meet: does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation*, **39**(1), 58–70.

Bocker, K.B., Baas, J.M., Kenemans, J.L., Verbaten, M.N. (2001). Stimulus-preceding negativity induced by fear: a manifestation of affective anticipation. *International Journal of Psychophysiology*, **43**(1), 77–90.

Bouton, M.E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, **52**(10), 976–86.

Bouton, M.E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, **11**(5), 485–94.

Bradley, M.M., Sabatinelli, D., Lang, P.J., Fitzsimmons, J.R., King, W., Desai, P. (2003). Activation of the visual cortex in motivated attention. *Behavioral Neuroscience*, **117**(2), 369.

Brockelmann, A.K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., Junghöfer, M. (2011). Emotion-associated tones attract enhanced attention at early auditory processing: magnetoencephalographic correlates. *The Journal of Neuroscience*, **31**(21), 7801–10.

Brunia, C.H., Hackley, S.A., van Boxtel, G.J., Kotani, Y., Ohgami, Y. (2011). Waiting to perceive: reward or punishment? *Clinical Neurophysiology*, **122**(5), 858–68.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**(1), 9–21.

Dolan, R.J., Heinze, H.J., Hurlemann, R., Hinrichs, H. (2006). Magnetoencephalography (MEG) determined temporal modulation of visual and auditory sensory processing in the context of classical conditioning to faces. *Neuroimage*, **32**(2), 778–89.

Duits, P., Cath, D.C., Lissek, S., *et al.* (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety*, **32**(4), 239–53.

Dye, L., Blundell, J.E. (1997). Menstrual cycle and appetite control: implications for weight regulation. *Human Reproduction*, **12**(6), 1142–51.

Ehlers, C.L., Hurst, S., Phillips, E., *et al.* (2006). Electrophysiological responses to affective stimuli in American Indians experiencing trauma with and without PTSD. *Annals of the New York Academy of Sciences*, **1071**(1), 125–36.

Fehm, L., Beesdo, K., Jacobi, F., Fiedler, A. (2008). Social anxiety disorder above and below the diagnostic threshold: prevalence, comorbidity and impairment in the general population. *Social Psychiatry and Psychiatric Epidemiology*, **43**(4), 257–65.

Felmingham, K.L., Bryant, R.A., Gordon, E. (2003). Processing angry and neutral faces in post-traumatic stress disorder: an event-related potentials study. *Neuroreport*, **14**(5), 777–80.

Flor, H., Birbaumer, N., Hermann, C., Ziegler, S., Patrick, C.J. (2002). Aversive Pavlovian conditioning in psychopaths: peripheral and central correlates. *Psychophysiology*, **39**(4), 505–18.

Flor, H., Birbaumer, N., Roberts, L.E., *et al.* (1996). Slow potentials, event-related potentials, 'gamma-band' activity, and motor responses during aversive conditioning in humans. *Experimental Brain Research*, **112**(2), 298–312.

Insel, T.R. (2014). The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *The American Journal of Psychiatry*, **171**(4), 395–7.

Junghöfer, M., Elbert, T., Tucker, D.M., Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, **37**(04), 523–32.

Keil, A., Müller, M.M., Gruber, T., Wienbruch, C., Stolarova, M., Elbert, T. (2001). Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clinical Neurophysiology*, **112**(11), 2057–68.

Keil, A., Stolarova, M., Moratti, S., Ray, W.J. (2007). Adaptation in human visual cortex as a mechanism for rapid discrimination of aversive stimuli. *Neuroimage*, **36**(2), 472–9.

Kluge, C., Bauer, M., Leff, A.P., Heinze, H.J., Dolan, R.J., Driver, J. (2011). Plasticity of human auditory-evoked fields induced by shock conditioning and contingency reversal. *Proceedings National Academy of Sciences of the United States of America*, **108**(30), 12545–12550.

Kotani, Y., Kishida, S., Hiraku, S., Suda, K., Ishii, M., Aihara, Y. (2003). Effects of information and reward on stimulus-preceding negativity prior to feedback stimuli. *Psychophysiology*, **40**(5), 818–26.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. (1997) Motivated attention: affect, activation, and action. In: *Attention and Orienting: Sensory and Motivational Processes*, pp. 97–135. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. (2005). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-6*, Gainesville, FL: University of Florida.

Lang, P.J., Davis, M., Ohman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of Affective Disorders*, **61**(3), 137–59.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, **23**, 155–84.

Lee, T.H., Lim, S.L., Lee, K.Y., Choi, J.S. (2009). Facilitation of visual processing by masked presentation of a conditioned facial stimulus. *Neuroreport*, **20**(8), 750–4.

Lissek, S., Powers, A.S., McClure, E.B., *et al.* (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Theraphy*, **43**(11), 1391–424.

Liu, Y., Keil, A., Ding, M. (2012). Effects of emotional conditioning on early visual processing: temporal dynamics revealed by ERP single-trial analysis. *Human Brain Mapping*, **33**(4), 909–19.

Lommen, M.J., Engelhard, I.M., Sijbrandij, M., van den Hout, M.A., Hermans, D. (2013). Pre-trauma individual differences in extinction learning predict posttraumatic stress. *Behaviour Research and Theraphy*, **51**(2), 63–7.

Lonsdorf, T.B., Menz, M.M., Andreatta, M., *et al.* (2017). Don't fear 'fear conditioning': methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews*, **77**, 247–85.

Lovibond, P.F., Liu, J.C., Weidemann, G., Mitchell, C.J. (2011). Awareness is necessary for differential trace and delay eyeblink conditioning in humans. *Biological Psychology*, **87**(3), 393–400.

Maris, E., Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, **164**(1), 177–90.

Michael, T., Blechert, J., Vriends, N., Margraf, J., Wilhelm, F.H. (2007). Fear conditioning in panic disorder: enhanced resistance to extinction. *Journal of Abnormal Psychology*, **116**(3), 612–7.

Michalowski, J.M., Melzig, C.A., Weike, A.I., Stockburger, J., Schupp, H.T., Hamm, A.O. (2009). Brain dynamics in spider-phobic individuals exposed to phobia-relevant and other emotional stimuli. *Emotion*, **9**(3), 306.

Miskovic, V., Keil, A. (2012). Acquired fears reflected in cortical sensory processing: a review of electrophysiological studies of human classical conditioning. *Psychophysiology*, **49**(9), 1230–41.

Moratti, S., Keil, A., Miller, G.A. (2006). Fear but not awareness predicts enhanced sensory processing in fear conditioning. *Psychophysiology*, **43**(2), 216–26.

Moratti, S., Keil, A., Stolarova, M. (2004). Motivated attention in emotional picture processing is reflected by activity modulation in cortical attention networks. *Neuroimage*, **21**(3), 954–64.

Mueller, E.M., Panitz, C., Hermann, C., Pizzagalli, D.A. (2014). Prefrontal oscillations during recall of conditioned and extinguished fear in humans. *Journal of Neuroscience*, **34**(21), 7059–66.

Mueller, E.M., Pizzagalli, D.A. (2016). One-year-old fear memories rapidly activate human fusiform gyrus. *Social Cognitive and Affective Neuroscience*, **11**(2), 308–16.

Ohgami, Y., Kotani, Y., Hiraku, S., Aihara, Y., Ishii, M. (2004). Effects of reward and stimulus modality on stimulus-preceding negativity. *Psychophysiology*, **41**(5), 729–38.

Panitz, C., Hermann, C., Mueller, E.M. (2015). Conditioned and extinguished fear modulate functional corticocardiac coupling in humans. *Psychophysiology*, **52**(10), 1351–60.

Pape, H.C., Pare, D. (2010). Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiological Reviews*, **90**(2), 419–63.

Peyk, P., De Cesarei, A., Junghöfer, M. (2011). ElectroMagnetoEncephalography Software: Overview and Integration with other EEG/MEG Toolboxes. *Computational Intelligence and Neuroscience*, **2011**, Article ID 861705, 10.

Peyk, P., Schupp, H.T., Elbert, T., Junghöfer, M. (2008). Emotion processing in the visual brain: a MEG analysis. *Brain Topography*, **20**(4), 205–15.

Peyk, P., Schupp, H.T., Keil, A., Elbert, T., Junghöfer, M. (2009). Parallel processing of affective visual stimuli. *Psychophysiology*, **46**(1), 200–8.

Pizzagalli, D.A., Greischar, L.L., Davidson, R.J. (2003). Spatio-temporal dynamics of brain mechanisms in aversive classical conditioning: high-density event-related potential and brain electrical tomography analyses. *Neuropsychologia*, **41**(2), 184–94.

Plochl, M., Ossandon, J.P., Konig, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, **6**, 278.

Poli, S., Sarlo, M., Bortoletto, M., Buodo, G., Palomba, D. (2007). Stimulus-preceding negativity and heart rate changes in anticipation of affective pictures. *International Journal of Psychophysiology*, **65**(1), 32–9.

Rothemund, Y., Ziegler, S., Hermann, C., *et al.* (2012). Fear conditioning in psychopaths: event-related potentials and peripheral measures. *Biological Psychology*, **90**(1), 50–9.

Santesso, D.L., Meuret, A.E., Hofmann, S.G., *et al.* (2008). Electrophysiological correlates of spatial orienting towards angry faces: a source localization study. *Neuropsychologia*, **46**(5), 1338–48.

Schirmer, A., Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. *Trends in Cognitive Sciences*, **21**(3), 216–28.

Schupp, H.T., Flaisch, T., Stockburger, J., Junghöfer, M. (2006a). Emotion and attention: event-related brain potential studies. *Progress in Brain Research*, **156**, 31–51.

Schupp, H.T., Ohman, A., Junghöfer, M., Weike, A.I., Stockburger, J., Hamm, A.O. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion*, **4**(2), 189–200.

Schupp, H.T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A.I., Hamm, A.O. (2006b). Stimulus novelty and emotion perception: the near absence of habituation in the visual cortex. *Neuroreport*, **17**(4), 365–9.

Sperl, M.F.J., Panitz, C., Rosso, I.M., *et al.* (2018). Fear extinction recall modulates human frontomedial theta and amygdala activity. *Cerebral Cortex*. 1–15.

Stein, M.B., Stein, D.J. (2008). Social anxiety disorder. *Lancet*, **371**(9618), 1115–25.

Steinberg, C., Dobel, C., Schupp, H.T., *et al.* (2012). Rapid and highly resolving: affective evaluation of olfactorily conditioned faces. *Journal of Cognitive Neuroscience*, **24**(1), 17–27.

Stolarova, M., Keil, A., Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: evidence for sensory plasticity in early affective perception. *Cerebral Cortex*, **16**(6), 876–87.

Van Strien, J.W., Franken, I.H., Huijding, J. (2014). Testing the snake-detection hypothesis: larger early posterior negativity in humans to pictures of snakes than to pictures of other reptiles spiders and slugs. *Frontiers in Human Neuroscience*, **8**, 691.

Tost, H., Champagne, F.A., Meyer-Lindenberg, A. (2015). Environmental influence in the brain, human welfare and mental health. *Nature Neuroscience*, **18**(10), 1421–31.

Vervliet, B., Craske, M.G., Hermans, D. (2013). Fear extinction and relapse: state of the art. *Annual Review of Clinical Psychology*, **9**, 215–48.

Walentowska, W., Paul, K., Severo, M.C., Moors, A., Pourtois, G. (2017). Relevance and uncertainty jointly influence reward anticipation at the level of the SPN ERP component. *International Journal of Psychophysiology*. **132**, Part B, 287–97. ISSN: 0167-8760.

Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews. Neuroscience*, **5**(4), 279–90.

Wessa, M., Flor, H. (2007). Failure of extinction of fear responses in posttraumatic stress disorder: evidence from second-order conditioning. *The American Journal of Psychiatry*, **164**(11), 1684–92.

Wiggert, N., Wilhelm, F.H., Boger, S., Georgii, C., Klimesch, W., Blechert, J. (2016). Social Pavlovian conditioning: short- and long-term effects and the role of anxiety and depressive symptoms. *Social Cognitive and Affective Neuroscience*, **12**(2), 329–39.

Wong, P.S., Bernat, E., Snodgrass, M., Shevrin, H. (2004). Event-related brain correlates of associative learning without awareness. *International Journal of Psychophysiology*, **53**(3), 217–31.

World Medical, A (2013). World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of American Medical Association*, **310**(20), 2191–4.

Yoon, S., Shim, M., Kim, H.S., Lee, S.-H. (2016). Enhanced early posterior negativity to fearful faces in patients with anxiety disorder. *Brain Topography*, **29**(2), 262–72.

Zuj, D.V., Palmer, M.A., Lommen, M.J., Felmingham, K.L. (2016). The centrality of fear extinction in linking risk factors to PTSD: A narrative review. *Neuroscience & Biobehavioral Reviews*, **69**, 15–35.