NeuroImage 205 (2020) 116302

Contents lists available at ScienceDirect

NeuroImage

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

Fear generalization of implicit conditioned facial features – Behavioral and magnetoencephalographic correlates

Kati Roesmann [a,b,*], Nele Wiens [a], Constantin Winker [a,b], Maimu Alissa Rehbein [a,b], Ida Wessing [a,b,c], Markus Junghoefer [a,b]

[a] Institute for Biomagnetism and Biosignalanalysis, University of Münster, Malmedyweg 15, D-48149, Münster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, D-48149, Münster, Germany
[c] Department of Child and Adolescent Psychiatry, University of Münster, Schmeddingstr. 50, D-48149, Münster, Germany

ARTICLE INFO

Keywords:
Fear generalization
Fear conditioning
Contingency awareness
Motivated attention
EEG
MEG

ABSTRACT

Acquired fear responses often generalize from conditioned stimuli (CS) towards perceptually similar, but harmless generalization stimuli (GS). Knowledge on similarities between CS and GS may be explicit or implicit. Employing behavioral measures and whole-head magnetoencephalography, we here investigated the neurocognitive mechanisms underpinning implicit fear generalization. Twenty-nine participants underwent a classical conditioning procedure in which 32 different faces were either paired with an aversive scream (16 CS+) or remained unpaired (16 CS-). CS+ and CS- faces systematically differed from each other regarding their ratio of eye distance and mouth width. High versus low values on this "threat-related feature (TF)" implicitly predicted the presence or absence of the aversive scream. In pre- and post-conditioning phases, all CS and 32 novel GS faces were presented. 16 GS+ faces shared the TF of the 16 CS+ faces, while 16 GS- faces shared the TF of the 16 CS- faces. Behavioral tests confirmed that participants were fully unaware of TF-US contingencies. CS+ compared to CS- faces revealed higher unpleasantness, arousal and US-expectancy ratings. A generalization of these behavioral fear responses to GS+ compared to GS- faces was observed by trend only. Source-estimations of event-related fields showed stronger neural responses to both CS+ and GS+ compared to CS- and GS- in anterior temporal (<100 ms) and temporo-occipital (<150 ms; 553–587 ms) ventral brain regions. Reverse effects were found in dorsal frontal areas (<100 ms; 173–203 ms; 257–290 ms). Neural data also revealed selectively enhanced responses to CS+ but not GS+ stimuli in occipital regions (110–167 ms; 330–413 ms), indicating perceptual discrimination. Our data suggest that the prioritized perceptual analysis of threat-associated conditioned faces in ventral networks rapidly generalizes to novel faces sharing threat-related features. This generalization process occurs in absence of contingency awareness and may thus contribute to implicit attentional biases. The coexisting perceptual discrimination suggests that fear generalization is not a mere consequence of insufficient stimulus discrimination but rather an active, integrative process.

1. Introduction

A quick differentiation of environmental cues predicting threat or safety is a key aspect of fear learning (Öhman and Mineka, 2001). To be adaptive, fear learning needs to be flexible such that stimuli similar to threat-related stimuli are considered potentially harmful as well. In social contexts, this ability to generalize fear (for review, see Dymond et al., 2015) may help to predict behaviors of others and to avoid negative interactions. However, an overgeneralization of defensive responses is considered a key feature and maintaining factor of anxiety disorders (Dunsmoor and Paz, 2015; Lissek et al., 2014a). Further, generalization processes may also play a key role in the evolution of social prejudices and race biases, which are thought to develop and change through processes akin to classical conditioning (De Houwer, Thomas and Baeyens, 2001; Olson and Fazio, 2006). In current political debates, for example, it is often argued that repeated reports on single immigrants who commit crimes might negatively affect attitudes and emotions towards people fleeing war and poverty. When the displayed criminals provide distinct facial features – in the following referred to as "threat-related features (TF)" – implicit and explicit generalized biases towards people sharing

* Corresponding author. Institute for Biomagnetism and Biosignalanalysis, University of Münster, Malmedyweg 15, D-48149, Münster, Germany.
E-mail address: k.roesmann@uni-muenster.de (K. Roesmann).

https://doi.org/10.1016/j.neuroimage.2019.116302
Received 28 March 2019; Received in revised form 19 September 2019; Accepted 17 October 2019
Available online 19 October 2019

these features might increase and influence various aspects of social interactions. According to classical conditioning approaches (Pavlov, 1927), such biases might arise because not only criminal individuals (i.e. conditioned stimuli, CS+), but also the TFs – shared with non-criminal individuals – might predict unconditioned aversive events (US, e.g. crimes).

Decades of research on animals and humans have revealed that aversively conditioned CS+ induce conditioned responses that resemble responses to the aversive US (Büchel and Dolan, 2000; Kim and Jung, 2006; Sehlmeyer et al., 2009). CS+ compared to the CS- (CS that is never paired with the US) typically induce stronger activations of the autonomic nervous system and differentially modulate neural responses in distributed brain structures of the fear-network (e.g. temporal lobe, prefrontal cortex; for reviews, see Büchel and Dolan, 2000; Sehlmeyer et al., 2009). Differential brain responses to CS+ compared to CS- emerge within the first 100 ms of stimulus processing (Bröckelmann et al., 2011; Hintze et al., 2014; Rehbein et al., 2014, 2015; Stolarova et al., 2006). This means that the brain differentiates threat-related from neutral stimuli very fast and presumably independent of explicit knowledge on CS/US contingencies (for review, see Miskovic and Keil, 2012).

Such implicit processes underlying fear acquision may be studied via MultiCS conditioning. MultiCS paradigms use a multitude of CS+ and CS- stimuli, e.g. different facial identities (Rehbein et al., 2014, 2015; Steinberg et al., 2012; for review, see Steinberg et al., 2013b). On a behavioral level, these studies yielded successful fear acquisition, evidenced by higher unpleasantness and arousal ratings of the CS+ compared to the CS-. Yet, the sensitivity measure d' in CS/US matching tasks suggested limited awareness in some MultiCS studies (for review, see Steinberg et al., 2013b; but see Junghöfer et al., 2016; Pastor et al., 2015). Thus, affective ratings might be partly independent from explicit knowledge on CS/US associations. On a neural level, MEG-based inverse source-reconstructions in MulitCS studies consistently indicated stronger activations to the CS+ compared to the CS- in right-lateralized frontal brain regions extending to anterior temporal sites (Rehbein et al., 2014; Steinberg et al., 2013a), and in temporo-occipital and inferior parietal (visual) brain regions (Steinberg et al., 2013a; Steinberg et al., 2012). Importantly, these differential effects started at short latencies (50–80 ms) after stimulus onset. Such affect-related extensive activation of the visual system was suggested to reflect "motivated attention", in which "appetitive or defensive motivational engagement directs attention and facilitates perceptual processing of survival relevant stimuli" (Bradley et al., 2003, p. 369; c.f. Schupp et al., 2004; Vuilleumier, 2005). Motivated attention was shown to modulate event-related potentials and fields (ERPs/ERFs) to various types of learned and biologically prepared affective stimuli at early (<130 ms, e.g. Keuper et al., 2014), mid-latency (~130–300 ms, Early Posterior Negativity, EPN; e.g. Junghoefer, et al, 2001; Keuper et al., 2014) and late processing stages (>300 ms, Late Positive Potential, LPP; e.g. Nelson et al., 2015; Pastor et al., 2015, for review, see Olofsson and Polich, 2007). ERP/ERF findings from MultiCS-conditioning studies (Steinberg et al., 2013b) and from studies presenting supra-versus subliminal affective stimuli (Keuper et al., 2018) suggest that particularly early and mid-latency emotion effects in visual brain regions might not depend on awareness of CS/US associations or on awareness of the affective content of the material, respectively. Thus, particularly early and mid-latency components seem to reflect relatively automatic perceptual responses to affective stimuli. However, it is still unknown, if such automatic responses may generalize to stimuli that have never been paired with a US, but that share specific features with the CS+ or CS-.

Lissek et al. (2008) were the first to study fear generalization in humans using classical conditioning. After a typical acquisition phase, presenting one CS+ (a large ring) and one CS- (a small ring), they introduced a generalization phase that presented 8 generalization stimuli (GS) on a perceptual continuum between CS+ and CS- (8 rings of different medium sizes) that had never been paired with the US. As a result, perceived risk ratings as well as fear-potentiated startle (FPS) responses were not only higher for the CS+ compared to the CS-, but also for GS similar to the CS+. More specifically, fear responses to the different GS increased with their increasing perceptual similarity with the CS+. Since this seminal study, effects of fear generalization on behavioral and neurophysiological autonomous correlates of fear (e.g. SCR, FPS) have been replicated using different US (e.g. electro shocks, loud noise), and different types of simple and complex visual CS and GS, including geometric shapes and colors (Lissek et al., 2010; Lissek et al., 2014a,b; McTeague et al., 2015; Vervliet and Geens, 2014; Vervliet et al., 2010), morphed faces (Haddad et al., 2012; Haddad et al., 2013; Onat and Büchel, 2015), and virtual rooms (Andreatta et al., 2015; for reviews, see Dunsmoor and Paz, 2015; Dymond et al., 2015).

During the last decade, a growing body of studies investigated the neural underpinnings of fear generalization (Dymond et al., 2015; Greenberg et al., 2013; Onat and Büchel, 2015). Yet, research on the role of "motivated attention" and explicit knowledge of CS/US contingencies in fear generalization is scarce. Some evidence for a generalization of "motivated attention" from the CS+ to GS comes from a parametric fMRI study (Lissek et al., 2014a) showing generalized, threat-associated neural responses in bilateral inferior parietal regions – i.e. in regions involved in processes of motivated attention (Bradley et al., 2003; Schupp et al., 2004; Vuilleumier, 2005). In line with that, a study using steady-state Visually Evoked Potentials (ssVEPs) (McTeague et al., 2015) showed generalization effects over parietal electrode sites – i.e. stronger ssVEPs for the CS+ and GS with higher similarity to the CS+. Interestingly, this was accompanied by highly discriminative, CS+-specific enhancements of ssVEPs over central occipital sites. The only ERP study investigating the role of motivated attention in fear generalization (Nelson et al., 2015) focused on the LPP. This late ERP component (>300 ms) has been associated with activity in temporo-parietal (visual) brain regions (Sabatinelli et al., 2007) and more elaborate attentional processes (Schupp et al., 2006). The expected generalization gradient, with higher LPP amplitudes for CS+ and GS with higher similarity to the CS+, was found at least in a subgroup of participants. This subgroup was characterized by *low* levels of "intolerance of uncertainty", a cognitive bias which – with high levels – promotes anxiety in uncertain situations (Dugas et al., 2004). On the other hand, participants with high "intolerance of uncertainty" showed higher LPP amplitudes in response to CS+ relative to perceptually similar GS. The authors speculated that the lack of generalization effects in these participants might be a consequence of secondary strategies, such as attentional avoidance, that were previously shown to influence LPP effects (MacNamara and Hajcak, 2009). Thus, the LPP may partly reflect a generalization of stimulus-driven "motivated attention", but this rather late event-related component may also be influenced by cognitive strategies based on explicit knowledge on CS/US contingencies.

Due to the lack of event-related electro- and magnetoencephalographic (EEG, MEG) studies on fear generalization focusing on earlier components, the role of more automatic and implicit stages of stimulus processing in fear generalization has remained unclear. In addition, all three studies revealing generalization effects in areas associated with motivated attention (Lissek et al., 2014a; McTeague et al., 2015; Nelson et al., 2015) used GS that differed from the CS+ in a rather obvious dimension (i.e. the orientation, the size of rings, the width of rectangles). Thus, it needs to be addressed whether generalization effects in visual areas rely on explicit information or whether they reflect automatic perceptual tuning processes (Olson and Fazio, 2006; Steinberg et al., 2012, 2013b)

Employing behavioral measures and whole-head high-density MEG, we here investigated the neurocognitive mechanisms of implicit fear generalization. In this context, the temporal dynamics of fear generalization on the one hand and CS/GS discrimination on the other were examined. In a MultiCS conditioning paradigm, faces of 32 different individuals were either paired with an aversive auditory US (16 CS+) or remained unpaired (16 CS-). CS+ and CS- faces systematically differed regarding their ratio of eye distance and mouth width – i.e. the threat-related feature (TF). In a following generalization phase, all 32 CS and

another 32 GS, i.e. additional faces sharing the TF+ of the CS+ (16 GS+) or the TF- of the CS- (16 GS-), were presented. Following MultiCS conditioning, the participants' TF/US contingency awareness (i.e. explicit knowledge on associations between the TF and the US) and CS/US contingency awareness (i.e. explicit knowledge on associations between individual CS faces and the US) were captured. Moreover, valence-, arousal-, and US-expectancy ratings in response to CS and GS faces were collected. Based on previous implicit conditioning (Olson and Fazio, 2006; Steinberg et al., 2013b) and generalization studies (Dymond et al., 2015; McTeague et al., 2015; Nelson et al., 2015), we expected stronger unpleasantness, arousal, and US-expectancy ratings of faces with the TF+ (i.e. CS+ and GS+) compared to faces with the TF- (CS- and GS-). We predicted differences between CS+ and CS- to be stronger than between GS+ and GS-, because the predicted GS+/GS- differentiation relies on implicit TF/US contingencies only. On a neural level, we hypothesized effects of motivated attention to faces with the TF + compared to the TF- as indicated by differential brain activation in visual brain areas (Lissek et al., 2014a; McTeague et al., 2015; Nelson et al., 2015), particularly in early and mid-latency components reflecting relatively automatic perceptual responses to affective stimuli. In addition to these effects of generalization, we predicted CS+-specific responses in the visual cortex indicating a successful discrimination of CS and GS (McTeague et al., 2015).

2. Methods

2.1. Participants

Twenty-nine healthy native German speakers (21 females) with a mean age of 24.90 years (SD = 5.91) participated in this study. All twenty-nine participants entered the behavioral analyses. For MEG analyses, the neural data of three female subjects were excluded because the maximum and the mean values of the global power of the relevant difference topographies (Generalization minus Baseline; see below) across the time interval of interest (0–600 ms) exceeded the median of all participants by more than three scaled median absolute deviations.[1] Participants had normal or corrected-to-normal visual acuity and reported no current or former severe neurological or psychiatric disorder. The sample was characterized by normal levels of anxiety (German version of the *State Trait Anxiety Inventory, STAI,* Laux et al., 1981), intolerance of uncertainty (German *Intolerance of Uncertainty* questionnaire, *UI-18,* Gerlach et al., 2008) and depression (short version of the German *General Depression Scale, ADS-K*, Hautzinger and Bailer, 1993; see Table 1). All participants provided written informed consent for their participation in this study and were compensated with 20 Euro or course credit.

3. Material

3.1. Conditioned and generalization stimuli (CS, GS)

One hundred male and female faces with a neutral facial expression were pre-selected from the *Radboud Faces* database (Langner et al., 2010) and the *Karolinska Directed Emotional Faces* database (KDEF; Lundqvist et al., 1998). A selection process served the goal to realize the four experimental cells of our 2 x 2 within-subject design (see Fig. 1B) with the factors Stimulus Type (CS, GS) and TF (TF+, TF-): We first excluded faces with pop-out effects (e.g. highly attractive or highly unattractive faces). Second, to realize the factor TF, we (1) measured the distance (in mm) from one pupil to the other (pupil distance), and the distance (in mm) from the left to the right corner of the mouth (mouth width) in each face, (2) we calculated the ratio of pupil distance and mouth width, and (3) we selected faces with either low or high ratios of pupil distance and mouth width (low-TF vs high-TF; examples of the male and female individuals with the lowest and the highest pupil distance to mouth ratio respectively are presented in Fig. 1A). Note that the ratio between pupil distance and mouth width naturally varied amongst individual faces. The ratio of pupil distance and mouth width was chosen as the TF because (1) differences in this measure between high-vs. low-TF faces were presumably not obvious (i.e. implicit[2]) without prior knowledge on the relevant TF, and because (2) this measure was clearly defined and measureable in each face of the two data bases.

Table 1
Description of the sample (N = 29).

	M	SD	Min	Max
Age	24.90	5.91	18	42
Years of Education	13.84	2.08	11	19
STAI (overall)	76.86	13.38	54	101
State	37.38	7.72	26	53
Trait	39.48	8.25	27	61
ADS-K	8.03	6.03	0	24
UI-18	42.14	12.42	22	70

M = mean; *SD* = standard deviation; Questionnaires were obtained to characterize the sample regarding their levels of state and trait anxiety (State Trait Anxiety Inventory , STAI), depression (ADS-K) and intolerance of uncertainty (UI-18), as these constructs were shown to influence processes of fear generalization (Greenberg et al., 2013; Nelson et al., 2015).

For the final stimulus selection, we assigned 32 male and 32 female faces to the 4 sets of stimuli and thereby balanced the amount of female and male faces in each set (8 male faces, 8 female faces; see Fig. 1B). Two sets of 16 stimuli each (TF-low-a and TF-low-b) scored equally low on the average ratio of pupil distance to mouth width ($t(30) = 0.19$, $p = .84$), while the other two sets (TF-high-a and TF-high-b) scored equally high on this ratio ($t(30) = 0.00$, $p = 1.00$). As intended, the average TF strongly differed between the TF-low and the TF-high sets ($t(30) = 13.32$, $p < .001$), but did not differ between male and female faces ($t(62) = 0.50$, $p = .62$). Descriptive statistics for the TFs in these selected faces are presented in Table 2. Using Adobe Photoshop® Version 11.0 (San Jose, California, USA), we scaled each face to a chin/vertex distance of 13.0 cm and a left-ear/right-ear distance of 10.3 cm (not changing the pupil distance to mouth width ratio). To minimize the influence of other potentially confounding perceptual factors, all faces were converted to gray scaled images, adjusted in brightness and presented centrally on an isoluminant gray background.

3.2. Unconditioned stimulus

The US was a loud, aversive female scream taken from the International affective sound system (Bradley and Lang, 1999) and shortened to 1200 ms containing the most emotionally arousing interval. The scream was presented at 60 dB above the individual hearing thresholds, which were determined individually for each subject's ear prior to the experimental tasks in the MEG. With an average loudness of 96.55 dB on the right ear and 92.24 dB on the left ear, the loudness was comparable with previous studies on MultiCS conditioning (Rehbein et al., 2015).

3.3. Stylized faces for the pair comparison task

The ability of participants to identify the relevant TF was tested by comparing the actually used TF+/TF- with seven "fake TF+/TF-". To this end, we used Adobe Photoshop© to create 8 pairs of stylized and gray-scaled faces (side by side). The right face of each pair differed from the left one regarding the ratio or size of two out of six facial features. These features included the eyes (pupil distance) and the mouth (mouth width),

[1] The scaled median absolute deviation is defined as K*MEDIAN(ABS(A-MEDIAN(A))) where K is the scaling factor and is approximately 1.48.

[2] In contrast to potentially obvious facial features such as length of the nose.

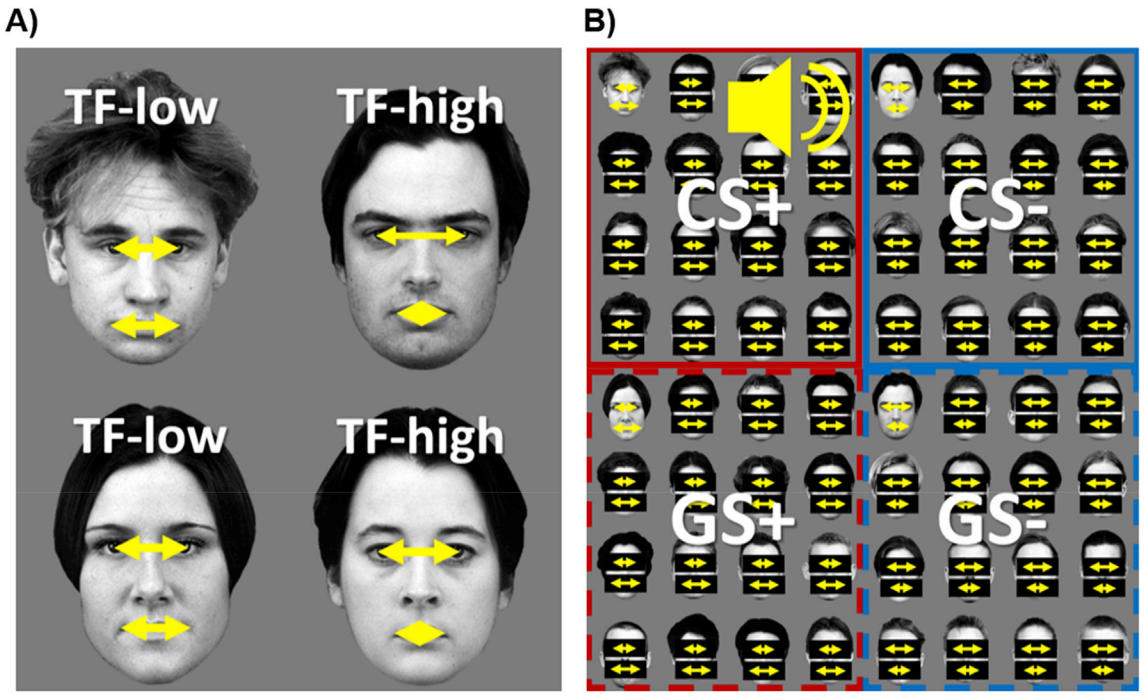

Fig. 1. A) Examples of male and female faces with a rather low and a rather high pupil distance to mouth width ratio, i.e. the threat-related feature (TF-low; TF-high). Pupil distance and mouth width are indicated by arrows. Faces depicted in this figure are adapted from the *Karolinska Directed Emotional Faces* database (AM32NES, AM01NES, AF30NES, AF14NES). B) Exemplified stimulus/condition assignment: 32 different faces with high pupil distance to mouth width ratio (TF-high) were assigned as CS+ and GS+ respectively (16 each). Another 32 faces with low pupil distance to mouth width ratio (TF-low) were assigned as CS- and GS- (16 each). Only CS+ faces predicted the US (loud scream) while CS-, GS+ and GS- faces were never paired with a US. The assignment of the four face sets to CS+/GS+ and CS-/GS- was counterbalanced across participants but CS+/GS+ and CS-/GS- sets always shared the same TF specification. Arrows and eye/mouth-masks are superimposed for illustration only and were not presented in the study.

Table 2
Stimulus characteristics and counterbalancing scheme.

		Threat-related feature (TF): (Pupil distance in mm)/Mouth width in mm)				Counterbalancing			
Set		*M*	*SD*	*Min*	*Max*	a ($N = 7$)	b ($N = 7$)	c ($N = 7$)	d ($N = 8$)
TF-low	a	1.10	.05	.97	1.16	CS+	GS+	CS–	GS–
	b	1.10	.04	1.04	1.16	GS+	CS+	GS–	CS–
TF-high	a	1.28	.07	1.21	1.44	CS–	GS–	CS+	GS+
	b	1.28	.06	1.22	1.38	GS–	CS–	GS+	CS+

Left: Characteristics of the TF (ratio: pupil distance in mm / mouth width in mm) within each of the four experimental sets (TF-low-a, TF-low-b, TF-high-a, TF-high-b). Each set consisted of eight female and eight male faces. *M* = mean, *SD* = standard deviation, *Min* = minimum, *Max* = maximum. Right: For counterbalancing purposes four experimental versions (a, b, c, d) were employed which differed regarding the assignments of sets to the 4 experimental conditions (CS+, GS+, GS–, CS–). *N* = number of participants receiving each of the 4 versions.

as well as the nose, the ear, the eyebrows, and the head shape. For example, the stylized face pair representing the actually used TF showed two stylized faces with different ratios of pupil distance to mouth width. The two features (TF or "fake TF") that distinguished the two faces of a pair were highlighted in red.

3.3.1. Procedure

All participants were tested at the Institute of Biomagnetism and Biosignalanalysis, University Hospital Münster. Ethical approval was obtained from the Ethics Committee of the Medical Faculty of the University of Münster. Ethical standards of the Declaration of Helsinki were strictly followed.

After giving informed consent, participants completed self-report questionnaires. The participants' head shapes were digitized using a 3D tracking device (Polhemus, Colchester, VT, USA; http://www.polhemus.com/). Participants were then seated in the MEG-chamber and landmark coils in each earlobe and on the nasion monitored their head position. The overall experimental procedure (see Fig. 2) adhered to methodological guidelines for fear conditioning research as described by Lonsdorf et al. (2017). Before the MEG measurements, the experiment started with baseline valence and arousal ratings of all stimuli (Baseline-Rating, details see below). Next, MEG was measured during the MultiCS paradigm (Baseline-MEG, Conditioning-MEG and Generalization-MEG). Upon completion of all MEG measurements, all stimuli were rated again (Generalization-Rating). Participants were then guided to the behavioral lab, in which three additional behavioral tasks were conducted in the following order: 1) CS/US Matching Task, 2) US-Expectancy Task, 3) Pair Comparison Task. All tasks were programmed using Presentation® Version 19.0. Overall, this experiment lasted for 2 h.

3.4. MultiCS paradigm

In all phases of the MultiCS paradigm, including Baseline-MEG,

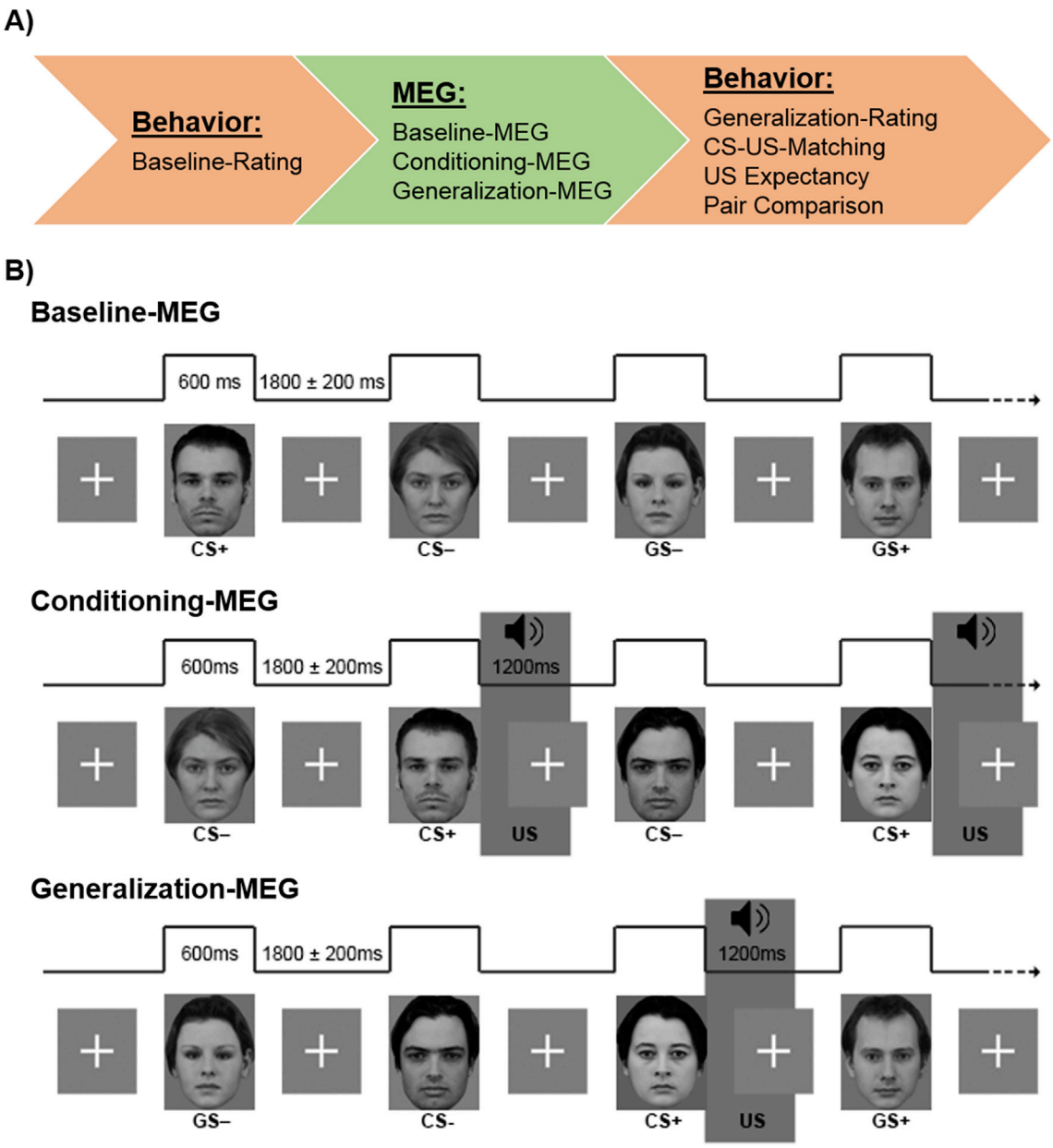

Fig. 2. Experimental procedure. A) Sequence of experimental tasks. B) MultiCS paradigm used in the MEG. In the Conditioning-MEG phase, multiple faces were either always paired (CS+) or remained always unpaired (CS-) with an auditory aversive scream (US). CS+ and CS- differed systematically in the relationship between pupil distance and mouth width (TF: TF + or TF-). In the Baseline-MEG phase before and the Generalization-MEG phase after Conditioning-MEG, CS+, CS- and additional generalization stimuli (GS) were presented. Thereby, half of the GS shared the TF of the CS+ (TF+), i.e. GS+ faces, whereas the other half shared the TF of the CS- (TF-), i.e. GS- faces. The assignment of the four face sets to CS+/GS+ and CS-/GS- was counterbalanced across participants. Faces depicted in this figure are adapted from the *Karolinska Directed Emotional Faces* database (AF09NES, AF14NES, AF022NES, AM13NES, AM01NES, AM29NES).

Conditioning-MEG and Generalization-MEG (Fig. 2), participants passively viewed the displayed faces without any task. Participants were informed that loud screams (US) would be presented in the conditioning and the generalization phase, but received no information regarding CS/US and TF/US contingencies. Faces were presented one by one in the center of the screen with a viewing angle of 8.26° × 6.55° (edge to edge). Each trial started with a fixation cross that was presented for 1800±200 ms and replaced by a face that was presented for 600 ms. In each phase, faces were presented 4 times in a pseudorandomized order with not more than three consecutive repetitions of the same condition (CS+, CS-, GS+ or GS-). Stimuli did not repeat until the complete set of 64 individual face stimuli had been presented once.

In the Conditioning-MEG phase, each CS+ face was always followed by the US, while CS- faces were never paired with the US (100% contingency). Overall, 128 trials were presented (16 face stimuli per set x 2 conditions (CS+, CS-) x 4 presentations). Because CS+ and CS- face sets differed with regard to the TF (low or high pupil to mouth width ratio), CS+/US associations were linked to both the face identity and the TF. The assignment of TF-low (a or b) and TF-high (a or b) faces to the CS- or CS+ condition was counterbalanced across participants.

The conditioning phase was preceded by a Baseline-MEG phase and followed by a Generalization-MEG phase. Both phases presented all face sets (CS+, CS-, GS+, GS-) and comprised 256 trials each (16 face stimuli per set x 4 conditions (CS+, CS-, GS+, GS-) x 4 repetitions). In the baseline phase, the US never appeared. In contrast, in the generalization phase, the CS+ continued to predict the US with 100% contingency. Importantly, the US was never paired with CS-, GS+ or GS- faces.

3.5. MEG recording and MEG data Preprocessing

During all three MEG phases, ERFs were acquired using a 275-sensor whole-head MEG system (Omega 275, CTF, VSM MedTech Ltd., Coquitlam, Canada) with first-order axial SQUID gradiometers. Continuous signals in a frequency range of 0 Hz–150 Hz were acquired with a sampling rate of 600 Hz. Neurophysiological data were further processed and analyzed with the Matlab-based *Electromagnetic Encephalography Software* EMEGS (Version 2.9, Peyk et al., 2011). Offline, data were first filtered using a 0.1 Hz zero-phase high-pass filter (forward/backward 2nd-order Butterworth-filter) and a 48 Hz low-pass filter (4th order) and then down-sampled to 300 Hz. Signals within epochs from 200 ms before to 600 ms after stimulus onset were baseline-adjusted for an interval from −150 ms to stimulus onset. Artifact detection and rejection was performed with an established method for the statistical control of artifacts in high-density electro- and magnetoencephalography data (Junghöfer et al., 2000). This procedure (1) detects individual sensor artifacts; (2) detects global artifacts; (3) replaces artifact-contaminated sensors by spherical spline interpolation statistically weighted on the basis of all remaining sensors; and (4) computes the variance of the signal across trials to document the stability of the averaged waveform. The rejection of artifact-contaminated trials and the interpolation of artifact-contaminated sensors relies on the calculation of statistical parameters for the absolute measured magnetic field amplitudes over time, their standard deviation over time, as well as on the determination of boundaries for each parameter based on their distribution across trials. If the goodness of test topography interpolations based on the residual sensor configuration within a given trial did not reach a predefined minimum criterion ($k = 0.01$; identical for each subject and run) the respective trial was rejected. On average 2.4 trials out of 64 trials (16 faces x 4 repetitions) per condition per person were rejected. A one-way ANOVA confirmed that each condition contained comparable mean numbers of trials ($F(3, 75) = 0.41$, $p = .748$; CS+: $M = 61.61$, $SD = 2.10$; CS-: $M = 61.61$, $SD = 2.23$; GS+: $M = 61.73$, $SD = 2.09$; GS-: $M = 61.46$, $SD = 2.25$). No subject was excluded because of a deviating number of rejected trials (>3 scaled median absolute deviation distance from the sample median). The signal was averaged across trials and separately for each participant, baseline and generalization phases, and each face set (CS+, CS-, GS+, GS-).

3.6. L2-minimum norm estimation (L2-MNE)

After averaging, we estimated the underlying neural sources of the recorded ERFs using the L2 minimum norm approach (L2-MNE, Hämäläinen and Ilmoniemi, 1994). This inverse modelling approach yields estimates of distributed cortical network activity without prior assumptions regarding the number and/or location of current sources (Hauk, 2004). The L2-MNE was based on a spherical head model with 350 evenly distributed dipole pairs (azimuthal and polar direction) and a source shell radius of 87% of the individually fitted head. The Tikhonov regularization parameter Lambda was set to 0.1. This resulted in topographies of neural activity for the baseline, conditioning and generalization phases. Note that the acquisition of conditioned fear responses as a function of repeated pairings of CS+ faces with the US during the acquisition phase is a presupposition for generalization processes. Therefore, analyses and results of the conditioning phase can be found in the supplementary materials S1. Examining neurocognitive processes of generalization and CS+/GS discrimination, this work focuses on neural responses in the generalization phase compared to baseline. To reduce variance due to perceptual stimulus characteristics of the individual faces employed as CS+, CS-, GS+, and GS-, difference topographies (Generalization – Baseline) were computed. For visualization purposes, L2-MNE results were projected on standard 3D brain models.

3.7. Statistical analysis of L2-MNE

The obtained difference topographies (Generalization – Baseline) were analyzed using a pointwise repeated-measures 2 x 2 ANOVA with the within-subject factors Stimulus Type (CS, GS) and TF (TF+, TF-). To test our hypothesis that implicitly acquired CS+/US associations of the TF+ would generalize from CS+ to GS+ faces, we first focused on main effects of the factor TF (TF+, TF-) as an index of generalization. Second, to explore differential effects of CS and GS, we additionally computed the interaction of TF and Stimulus Type as an index of discrimination. To account for multiple testing of the point-wise ANOVA, we applied a non-parametric statistical testing approach (Maris and Oostenveld, 2007). As part of this procedure, so-called clustermasses within the time window of interest (0–600 ms after stimulus onset) were calculated. Clustermasses represent the sum of F-values of test-dipoles that revealed significant effects at $\alpha = .05$ (sensor-level criterion) in at least five spatially neighboring dipoles and five temporally consecutive time points. Observed clustermasses were then tested against a distribution of the maxima of random clustermasses that were generated via Monte-Carlo simulations of identical analyses based on 1000 random permutations of the experimental conditions. Only clustermasses that exceeded an alpha level of $\alpha = .05$ (cluster-level criterion) within the test interval from 0 to 600 ms were reported. Bonferroni-corrected post-hoc t-tests (critical p-value: $p/6 = .05/6 = .008$) were computed within all clusters revealing significant interaction effects of Stimulus Type (CS vs. GS) and TF (TF+ vs. TF-).

3.8. Behavioral tasks

3.8.1. Valence and arousal ratings

We used a computerized version of the Self-Assessment-Manikin (SAM) scale (Bradley and Lang, 1994) to obtain ratings of valence and arousal for all faces. Each dimension was rated on a scale ranging from 1 to 9 by mouse clicks, with higher numbers indicating higher pleasantness or arousal, respectively. Ratings of all 64 faces were obtained directly before (Baseline-Rating) and after (Generalization-Rating) the MEG measurement. For each rating dimension (valence, arousal) and condition (CS+, CS-, GS+, GS-) difference scores (Generalization-Rating – Baseline-Rating) were computed. First, we tested the hypothesis that stimuli sharing the TF+ (CS+, GS+) were rated as more unpleasant and arousing after conditioning by means of Bonferroni-corrected (one--sample t-tests (one-sided, critical p-value: $p/4 = .05/4 = .0125$). In particular, difference scores (Generalization-Rating – Baseline-Rating) of CS+, CS-, GS+ and GS- were compared with 0. Second, difference scores were analyzed using within-subject 2 x 2 ANOVAs with the factors Stimulus Type (CS, GS) and TF (TF+, TF-). Significant interaction effects were further delineated using Bonferroni-corrected paired t-tests. Based on the expectation of higher unpleasantness and arousal ratings specifically of the CS+ and (possibly to a lesser extend) the GS+, one-sided t-tests were used, if applicable (one-sided: CS+ >= GS+, CS+ > CS-, CS+ > GS-, GS+ > CS-, GS+ > GS-, two-sided: CS- vs. GS-, critical p-value: $p/6 = .05/6 = .008$).

3.9. US-expectancy rating

In the US-expectancy rating, participants indicated for all faces (CS+, CS-, GS+, GS-) whether they would expect a scream after its presentation. Ratings were obtained via mouse click on a computerized 8-point Likert scale ranging from 0% (“I am sure there will be no scream”) to 100% (“I am sure there will be a scream”). To maintain the expectancy of US presentations and prevent extinction within this task, each rating of a CS+ face was followed by a scream, while ratings of CS-, GS+ and GS- faces were not. Data were analyzed using a within-subject 2 x 2 ANOVA with the factor Stimulus Type (CS, GS) and the factor TF (TF+, TF-). If applicable, one-sided Bonferroni-corrected paired t-tests (see valence and arousal ratings) were used to further delineate interactions.

3.10. CS/US matching task

The CS/US matching task measuring the level of CS/US contingency awareness was validated by and adopted from Rehbein and colleagues (Rehbein et al., 2015). For each CS+ and CS- face (but not for the GS), participants indicated by mouse click whether it was paired with the US by means of an 8-point Likert scale ranging from −4 ("There was definitely no scream") to 4 ("There was definitely a scream"). Thereby, negative and positive values indexed an assignment to the CS- and CS+ condition, respectively. The level of awareness concerning the stimulus category was assessed by means of the sensitivity measure d' (Wickens, 2002). The absolute value of the response served as a measure of certainty (1 = weak certainty to 4 = strong certainty). Subjective certainties of negative (i.e. correct rejections, misses) and positive (i.e. hits, false alarms) responses were further investigated using two separate Wilcoxon signed rank tests.

3.11. Pair comparison task

The pair comparison task was employed to investigate individual levels of TF/US contingency awareness. It started with a partial debriefing, informing the participants that the presentation of the US was not random but associated with certain facial features. In a forced-choice pair comparison task, participants were instructed to indicate by a left or right mouse click, whether they considered the distinctive features of the left or of the right pair of stylized faces more likely to predict the US. In each trial of this task, participants were confronted with two pairs of stylized faces, which were presented on the left and the right side of the screen. For example, stylized faces of the left pair might differ from each other regarding the ratio of pupil distance to mouth width, i.e. the actually used TF. Stylized faces of the right pair might differ regarding the size of the eyes relative to the vertical placement of the ears, i.e. a "fake TF". Upon participants' selection of one pair, the next trial started. Each of the eight stylized face pairs was presented once with any of the other seven stylized face pairs, resulting in overall 28 comparisons. Thereby, the order of comparisons as well as the assignment of pairs to the right or left side of the screen was completely random and differed across participants. In the end of the task, participants were asked 1) whether they knew the correct answer and consistently chose the corresponding pair of stylized faces, or 2) whether they just guessed, or 3) if they followed a certain response pattern. Responses were documented by the experimenter.

To evaluate participants' TF/US contingency awareness, the frequency of choices for each pair of stylized faces was analyzed using frequency rankings and Laplace-probabilities for each pair. Indicated by the probability function of the binomial distribution, in cases of $n = 7$ comparisons of each pair with other pairs and a guessing probability $p = .5$, $k = 3$ correct answers are expected at chance level. According to the cumulative distribution function, the likelihood of $k > 5$ choices of a given pair is significant at a level of $\alpha < 0.05$. Thus, participants with $k >= 5$ choices of the actually used TF were defined as aware of TF/US associations, if they additionally indicated that they knew the correct answer and consistently chose the corresponding pair of stylized faces. In this case this participant would be excluded from further analyses.

4. Results

4.1. Behavioral measures of generalization and contingency awareness

4.1.1. Valence and arousal ratings

One-sample t-tests on the valence and arousal difference scores (Generalization – Baseline) revealed that both CS+ and GS+ were evaluated as more unpleasant (CS+: $t(28) = 3.61$, $p < .001$, $d = 0.67$; GS+: $t(28) = 1.88$, $p = .035$, $d = 0.35$) and more arousing (CS+: $t(28) = 3.87$, $p < .001$, $d = 0.72$; GS+: $t(28) = 1.96$, $p = .030$, $d = 0.36$) after compared to before the conditioning procedure. Effects for GS+ stimuli need to be interpreted with caution, as they lost significance after Bonferroni correction (critical p-value: $p/4 = .05/4 = .0125$). One-sample t-tests on the respective difference scores for CS- and GS- were non-significant, indicating that valence (CS-: $t(28) = 0.45$, $p = .653$, $d = 0.08$; GS-: $t(28) = 1.18$, $p = .245$, $d = 0.22$) and arousal (CS-: $t(28) = 0.67$, $p = .512$, $d = 0.12$; GS-: $t(28) = 1.07$, $p = .294$, $d = 0.20$) ratings in these conditions did not change as a function of learning.

In the valence ratings, the 2 x 2 ANOVA on the difference scores (Generalization – Baseline, Fig. 3 left) revealed a main effect of the factor TF ($F(1,28) = 7.48$, $p = .005$, $\eta^2 = 0.211$), with overall more unpleasant ratings of faces with the TF+ compared to faces with the TF-. As indicated by the significant interaction of Stimulus Type x TF ($F(1,28) = 4.72$, $p = .038$, $\eta^2 = 0.144$), this effect was mainly driven by more unpleasant ratings of the CS+ compared to the CS- ($t(28) = 3.12$; $p = .002$, $d = 0.62$), while more unpleasant ratings of the GS+ compared to the GS- were non-significant ($t(28) = 0.75$, $p = .23$, $d = 0.431$). CS+ elicited more unpleasant ratings than GS+ faces ($t(28) = 2.49$, $p = .010$, $d = 0.40$), while CS- and GS- faces were rated similarly ($t(28) = 0.95$, $p = .352$, $d = 0.13$). Indicating fear generalization, GS+ faces were rated as more unpleasant compared to CS- faces ($t(28) = 1.71$, $p = .049$, $d = 0.30$). After Bonferroni correction, however, differences between GS+ and CS-, but also differences between CS+ and GS+ lost significance (critical p-value: $p/6 = .05/6 = .008$).

Difference scores of arousal ratings (Generalization – Baseline, Fig. 3 middle) revealed a similar pattern. Again, we observed a main effect of the factor TF ($F(1,28) = 15{,}40$, $p < .001$, $\eta^2 = 0.355$). TF+ compared to TF- faces were rated as more arousing. Further, a significant interaction of Stimulus Type x TF ($F(1,28) = 8.72$, $p = .006$, $\eta^2 = 0.238$) indicated that this main effect was mainly driven by higher arousal ratings of CS+ compared to CS- ($t(28) = 4.88$, $p < .001$, $d = 0.61$). Higher arousal ratings of the GS+ compared to the GS- were not significant, but followed a trend in the hypothesized direction ($t(28) = 1.43$, $p = .083$, $d = 0.18$). CS+ elicited higher arousal ratings than GS+ ($t(28) = 2.31$, $p = .014$, $d = 0.36$), while CS- and GS- arousal ratings did not differ ($t(28) = 0.82$, $p = .418$, $d = 0.01$). GS+ faces were rated as more arousing compared to CS- ($t(28) = 2.01$, $p = .027$, $d = 0.244$). Again, after Bonferroni correction, differences between GS+ and CS-, marginal differences between GS+ and GS-, but also differences between CS+ and GS+ lost significance (critical p-value: $p/6 = .05/6 = .008$).

4.2. US-expectancy ratings

Like in valence and arousal ratings, the 2 x 2 ANOVA on US-expectancy ratings revealed a significant main effect of the factor TF ($F(1,28) = 15.54$, $p < .001$, $\eta^2 = 0.357$) with higher US-expectancy ratings for TF+ compared to TF- faces (Fig. 3 right). We further observed a significant main effect of the factor Stimulus Type ($F(1,28) = 17.22$, $p < .001$, $\eta^2 = 0.381$) with higher US-expectancy ratings of CS compared to GS. A significant interaction of Stimulus Type x TF ($F(1,28) = 11.64$, $p < .001$, $\eta^2 = 0.294$) indicated that both main effects were mainly driven by higher US-expectancy ratings in response to the CS+ compared to the CS- ($t(28) = 4.11$, $p < .001$, $d = 1.19$) and the GS+ ($t(28) = 4.14$, $p < .001$, $d = 1.20$), respectively. In tendency, GS+ induced the predicted higher US-expectancy ratings than GS- ($t(28) = 1.44$, $p = .087$, $d = 0.21$, non-significant after Bonferroni correction), while no difference was observed between GS+ and CS- ($t(28) = 0.01$, $p = .49$, $d = 0.00$)

4.3. CS/US-matching task

The sensitivity measure d' for the stimulus category (CS+ vs CS-) differed significantly from zero ($M = 1.381$, $SD = 0.998$, $t(28) = 7.46$, $p < .001$, 95% Confidence Interval [CI] = 1.00, 1.76), indicating that participants acquired at least partial CS+/US contingency awareness for individual CS faces. The Wilcoxon signed rank test indicated that certainty ratings (1–4) were significantly higher for correct vs. incorrect positive responses ("There was definitely a scream", hits: $M = 2.557$,

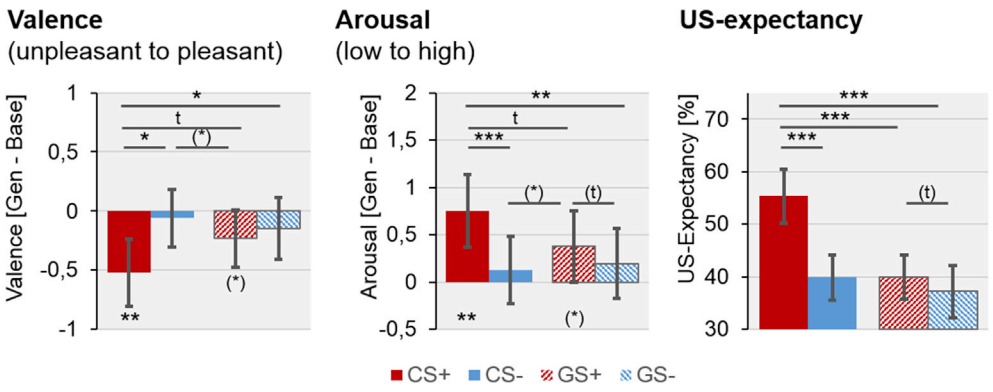

Fig. 3. Behavioral measures of generalization. Left: Mean difference scores (Generalization (Gen) minus Baseline (Base)) of valence ratings in response to CS+, CS-, GS+ and GS-. Negative values indicate more unpleasant valence ratings after versus before conditioning. Middle: Mean difference scores (Gen minus Base) of arousal ratings. Positive values indicate higher arousal ratings after versus before conditioning. Right: Mean US-expectancy ratings in %. Error bars indicate 95% confidence intervals of the means. The Bonferroni-corrected significance level of differences between conditions (above) and of changes within conditions (Gen minus Base; below) is indicated (*$p < .05$, **$p < .01$, ***$p < .001$, ^{t} $p < .1$, effects that did not survive Bonferroni correction are indicated in parentheses).

$SD = 0.687$, false alarms: $M = 1.812$, $SD = 0.788$, $Z = 3.781$, $p < .001$). For negative responses ("There was definitely no scream"), certainty ratings did not differ between correct rejections ($M = -2.225$, $SD = 0.564$) and misses ($M = -2.386$, $SD = 0.630$, $Z = 1.117$, $p = .264$).

4.4. Pair comparison task

The pair of stylized faces showing the actually used TF (pupil distance/mouth width) was selected in $M = 2.79$ ($SD = 1.30$) out of 7 pair comparisons, i.e. at chance level (k = 3). Mean choices of other pairs all ranged between $M = 2.52$ and $M = 4.24$. Each of the eight observed means significantly remained below the critical frequency of 6 ($t(28) = 4.96$ to 13.43, all p's < .001, critical p-value: $p = .05/8 = .006$), revealing that, on a group level, participants did not – correctly or incorrectly – link any of the depicted features to the US. To further explore whether individual participants reported contingencies between the relevant TF and US presentations, we looked at each mean separately. One participant correctly selected the relevant TF in 6 out of 7 comparisons i.e. significantly above chance level. However, in the post questionnaire this participant stated that she just guessed. Consequently, both group statistics and individual investigations suggest that participants were fully unaware of contingencies between the actually used TF and the US.

Behavioral findings in the subset of participants (N = 26) that entered the following MEG analyses are qualitatively identical to the results presented above (see supplementary materials S2).

4.5. MEG correlates of generalization (main effect: TF)

4.5.1. Ventral clusters with higher neural activation to faces with the TF+ compared to the TF.

The cluster-based permutation test for the main effect of TF revealed three significant clusters with relatively increased neural responses to faces with the TF+ compared to those with the TF- (red clusters in Fig. 4). While the first two clusters occurred at early latencies (Cluster V1: 67–90 ms, $F(1,25) = 30.13$, $p < .001$, $\eta^2 = 0.547$; Cluster V2: 103–147 ms, $F(1,25) = 16.42$, $p < .001$, $\eta^2 = 0.396$), the third cluster V3 was found at late latencies shortly before face offset (553–587 ms, $F(1,25) = 30.55$, $p < .001$, $\eta^2 = 0.550$). Cluster V1 spanned from right anterior temporal to inferior frontal regions. Cluster V2 was found in left occipito-temporal brain regions. Note that a cluster with very similar spatiotemporal characteristics (90–130 ms, left occipito-temporal) revealed an emerging differentiation of CS+ and CS- as a function of learning in the conditioning phase (see supplementary materials S1, Fig. S1). Similar to Cluster V2, Cluster V3 was located in left occipito-temporal brain areas, now extending to parietal regions. Thus, all three clusters were found in ventral brain regions of the visual stream. Importantly, and in contrast to the behavioral measures, in all three clusters the Stimulus Type x TF interactions were non-significant, showing that the main effect of TF was not mainly driven by differential CS+ and CS- processing but to a similar extent also by differential GS+ and GS- processing.

4.5.2. Dorsal clusters with lower neural activation to faces with the TF+ compared to the TF.

The permutation test for the main effect of TF revealed three additional significant clusters with relatively reduced brain activations in response to faces with the TF+ compared to those with the TF- (blue clusters in Fig. 4). The first cluster D1 was found at an early latency (43–84 ms, $F(1,25) = 36.91$, $p < .001$, $\eta^2 = 0.596$), while the second and third clusters were found at mid-latency time intervals (Cluster D2: 173–203 ms, $F(1,25) = 17.40$, $p < .001$, $\eta^2 = 0.410$; Cluster D3: 257–290 ms, $F(1,25) = 18.67$, $p < .001$, $\eta^2 = 0.428$). Cluster D1 occurred at left fronto-parietal regions. Cluster D2 and D3 also spanned left anterior parietal regions and extended to central prefrontal brain regions. Thus, all three clusters were found in dorsal brain regions. Again, in all three clusters the Stimulus Type x TF interactions were non-significant. Thus, differential CS+/CS- and GS+/GS- processing yielded similar effects.

4.6. MEG correlates of CS/GS discrimination (interaction: TF x Stimulus Type)

To further explore the neurocognitive basis of the observed behavioral interaction effects, with pronounced unpleasantness, arousal, and US expectancy ratings that were rather specific to the CS+, we conducted a second analysis which focused on spatiotemporal clusters revealing interactions of TF and Stimulus Type. We expected to replicate McTeague et al. (2015) who – in addition to generalization effects - showed clear signs of CS+/GS differentiation at occipital sites.

The cluster-based permutation test on these interactions in fact revealed two significant clusters in occipital brain regions (black clusters in Fig. 5). The first cluster occurred in central occipital sites between 110 ms and 167 ms (Cluster O1: $F(1,25) = 13.11$, $p < .001$, $\eta^2 = 0.344$). In this cluster, we found stronger brain activation in response to the CS+ compared to the CS- ($t(25) = 3.64$, $p = .001$, $d = 0.52$) and – marginally – to the GS+ ($t(25) = 2.84$, $p = .009$, $d = 0.37$, non-significant after Bonferroni correction). Other t-tests were all non-significant after Bonferroni correction (critical p-value: $p/6 = .05/6 = .008$): CS+ vs. GS- ($t(25) = 0.52$, $p = .61$, $d = 0.07$; CS- vs. GS+: ($t(25) = 0.68$, $p = .50$, $d = 0.10$; CS- vs. GS-: $t(25) = 2.49$, $p = .02$, $d = 0.45$, GS+ vs. GS-: $t(25) = 1.90$, $p = .07$, $d = 0.31$).

The second cluster extended towards more dorsal occipital sites at rather late latencies (Cluster O2: 330–413 ms, $F(1,25) = 18.08$, $p < .001$, $\eta^2 = 0.420$). In this cluster, CS+ again elicited stronger neural responses than CS- ($t(25) = 3.67$, $p = .001$, $d = 0.47$) and GS+ ($t(25) = 3.89$,

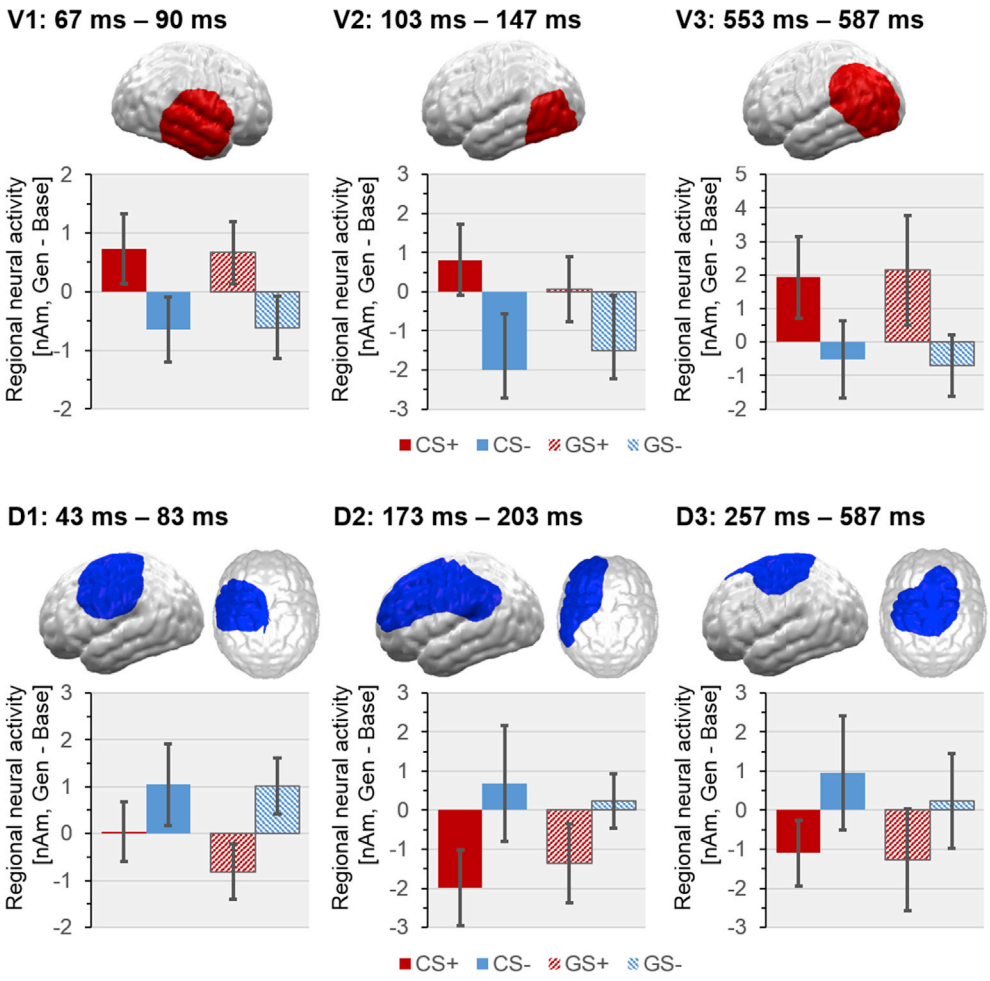

Fig. 4. Neural correlates of fear generalization indicated by the main effect of the threat-related feature (TF + vs TF-). Significant spatiotemporal clusters of the main effect TF were projected onto a 3D standard brain for visualization purposes. Top: Red clusters in ventral regions (V1, V2, V3) reveal higher neural activations to faces with the TF + compared to those with the TF-. Bottom: Blue clusters in dorsal areas (D1, D2, D3) show the reverse effect. Bar plots depict the regional mean neural activity (Generalization minus Baseline) within the respective clusters. Note that, within these clusters, TF x Stimulus Type interactions were all non-significant indicating undistinguishable differential effects of CS+/CS- and GS+/GS- processing. Error bars indicate 95% confidence intervals of the means.

$p = .001$, $d = 0.63$). Additionally, we found marginally lower responses to GS+ compared to GS- ($t(25) = -1.68$, $p = .016$, $d = 0{,}19$, non-significant after Bonferroni correction), while all other t-tests were non-significant (CS+ vs. GS-: $t(25) = 1.09$, $p = .29$, $d = 0.18$; CS- vs. GS+: $t(25) = 0.67$, $p = .50$, $d = 0{,}11$; CS- vs. GS-: $t(25) = 1.68$, $p = .105$, $d = 0.30$).

Importantly, both clusters revealed a clear discrimination not only between CS+ and CS- faces, but also between CS+ and GS+ faces. Thus, in addition to generalization of neural responses from CS+ to GS+ (Clusters V1-3, D1-3), we also found evidence for a sharp discrimination of CS+ and GS+ stimuli in the occipital cortex.

5. Discussion

We set out to investigate the neurocognitive mechanisms of implicit fear generalization based on conditioned facial threat-related features (TF) by means of a MultiCS conditioning paradigm. As intended, participants were unable to identify the relevant TF in the pair comparison task, rendering TF+/US associations fully implicit. However, in the CS/US matching task, participants associated CS and US above chance level, indicating successful learning of facial identities predicting the US. In addition, our supplementary analyses of the conditioning phase (see supplementary materials S1) revealed an emerging amplification of neural responses to the CS+ compared to the CS- as a function of learning. Confirming previous electrophysiological studies suggesting acquired fears to alter sensory processing (Miskovic and Keil, 2012), this effect was localized in left temporo-occipital, i.e. ventral brain regions (90–130 ms, see Fig. S1). Supporting our main hypothesis of implicit fear generalization based on the TF, we observed a main effect of the factor TF in the generalization phase, i.e. stronger fear-associated behavioral and neural responses in sensory brain regions to all stimuli with the TF+ (CS+, GS+) compared to those with the TF- (CS-, GS-). Beyond these predicted effects, dorsal brain regions revealed reduced brain activity for CS+ compared to CS- and for GS+ compared to GS-. Particularly, the observed GS+/GS- differentiations in ventral and dorsal brain regions in combination with similar tendencies in arousal and US-expectancy ratings provide support for implicit generalization mechanisms based on an implicit representation of the threat-related feature (TF). We also observed interactions between Stimulus Type and CS on the behavioral and on the neural level. As predicted, behavioral ratings (valence, arousal, US expectancy), and stimulus-induced activity in the occipital cortex indicated a successful discrimination of CS+ and GS+ stimuli (McTeague et al., 2015). In the light of results from the CS/US matching task, this finding might (partly) be influenced by explicit knowledge on CS+/US contingencies. In the following, we will discuss our findings in the light of the literature and in light of existing models of fear generalization.

5.1. Behavioral effects: generalization and discrimination

In line with the expected generalization of fear, not only the CS+, but – in tendency – also the GS+ were evaluated as more unpleasant and more arousing after versus before conditioning. By contrast, CS- and GS- ratings did not change as a function of learning. Further, faces with the TF+ compared to faces with the TF- revealed higher unpleasantness and arousal ratings as well as higher expectancy ratings. However, significant Stimulus Type x TF interactions indicated that this was mainly driven by a strong CS+/CS- differentiation, while the GS+/GS- differentiation was

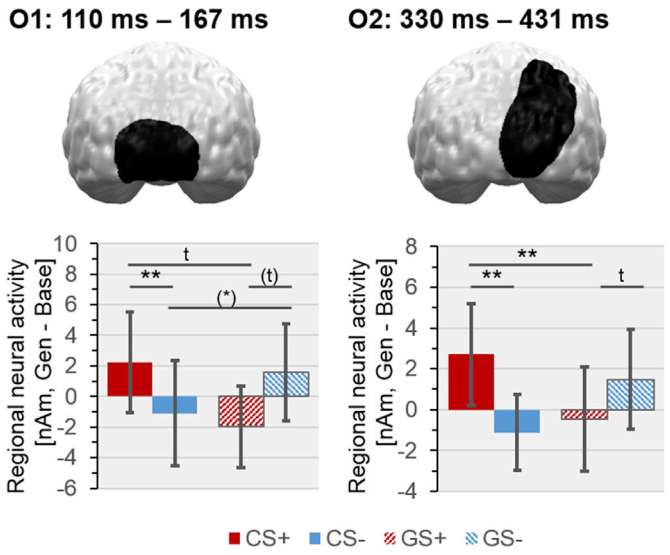

Fig. 5. Neural correlates of CS/GS discrimination indicated by the interaction effect of the TF (TF+ vs TF-) and the Stimulus Type (CS vs GS). Significant spatiotemporal clusters of the interaction effect of TF and Stimulus Type were projected onto a 3D standard brain for visualization purposes. Black clusters in occipital regions (O1 and O2) reveal higher neural activations to CS+ compared to CS- and GS+ faces. Bar plots depict the regional mean neural activity (Generalization minus Baseline) within the respective clusters. Error bars indicate 95% confidence intervals of the means. To further characterize interaction effects, we employed 6 paired-sample t-tests. The Bonferroni-corrected significance level of differences is indicated (*$p < .05$, **$p < .01$, ***$p < .001$, $^{t}p < .1$, effects that did not survive Bonferroni correction are indicated in parentheses).

only marginal in the arousal and US-expectancy ratings. The affective quality of the GS+ was rated in between the CS+ and CS- in all three behavioral outcome measures. On the one hand, GS+ were rated as more unpleasant and arousing compared to the CS-, pointing towards fear generalization. On the other hand, GS+ were also rated as less unpleasant and arousing compared to the CS+, pointing towards the ability of participants to discriminate GS+ from CS+. After Bonferroni correction, only differences between CS+ and both CS- and GS- remained significant, but not the discrimination of CS+ and GS+. Due to only marginal significance levels, behavioral evidence for generalization should be interpreted with caution. Based on these findings, we conclude that implicit behavioral generalization effects might be rather weak. This is in line with previous findings of behavioral conditioning effects under conditions of limited contingency awareness, which have been reported in some (Bröckelmann et al., 2011; Rehbein et al., 2014; Steinberg et al., 2012, 2013), but not all studies (Tabbert et al., 2011; Tabbert et al., 2006).

5.2. Ventral effects: enhanced motivated attention during fear generalization

Our study suggests that early-starting effects of motivated attention in ventral brain regions underpin fear generalization based on implicit facial features. The observation of generalization effects in occipito-temporal and inferior parietal, attention-related brain structures converges with generalization studies employing fMRI and ssVEP (Lissek et al., 2014a; McTeague et al., 2015). Further, our finding of late generalization effects (553–587 ms) in occipito-temporal and inferior parietal brain regions converges with Nelson et al.'s (2015) finding that fear generalization modulates the late positive potential (LPP), at least in participants with low but not high "intolerance of uncertainty". These authors suggested that the absence of generalization in highly intolerant participants may result from secondary strategies, such as attentional avoidance, presumably (partly) based on explicit knowledge on the dimension of generalization. The lack of TF+/US contingency awareness in our study might have prevented such deliberate avoidance, facilitating the finding of a late generalization effect in our sample.

Beyond these late effects of generalization, the current study showed that also fast neural retuning processes in sensory systems extend from conditioned stimuli to novel stimuli (i.e. GS+) that have never been paired with aversive events, but that share subtle perceptual features (i.e. the TF+) with CS+. In line with previous MultiCS studies (Rehbein et al., 2014; Steinberg et al., 2013a), we found enhanced neural responses to faces with the TF+ compared to those with the TF- in an anterior temporal and inferior frontal cluster as early as 67–90 ms after stimulus onset. Further, we found evidence for a mid-latency occipito-temporal differentiation (103–147 ms) for stimuli with the TF+ compared to those with the TF-, resembling the previously described late effect (553–587 ms). Interestingly, this sequence of effects converges with results of a previous MultiCS study, showing a transfer of conditioned responses from the originally conditioned CS+/CS- faces to same-identity faces with different expressions during similar processing stages (64–96 ms, 68–92 ms, 108–160 ms, 412–452 ms) in the EEG (Rehbein et al., 2018). Importantly, the spatiotemporal activation pattern observed here may indicate that late sustained attentional processes to generalization stimuli (e.g., Nelson et al., 2015) are initiated already at early perceptual processing stages. The recurrence of effects with a similar topography at different temporal processing stages (see also Steinberg et al., 2012, 2013) is in line with multiple-wave-models of affective processing (Pessoa and Adolphs, 2010), which assume that affective visual information is activated in 'multiple waves' that initiate and refine neural responses at a given processing stage. Overall, the spatio-temporal characteristics of the observed effects fit with previous research showing prioritized perceptual processing of motivationally relevant stimuli (Bradley et al., 2003; Schupp et al., 2004; Vuilleumier, 2005) and highlight that fast and automatic bottom-up processes indexing motivated attention play a key role in fear generalization.

5.3. Dorsal effects: reduced top-down control during fear generalization

Interestingly, *reduced* neural responses to stimuli with the TF+ compared to those with the TF- in dorsal fronto-parietal brain regions also emerged already at very early processing stages (43–83 ms) – i.e. temporally overlapping with ventral effects described above. Similar effects with dorsal topographies extending to central prefrontal brain regions again showed up at mid-latency time intervals (173–203 ms, 257–290 ms). We can only speculate on the functional significance of these effects. An influential model of attentional control in the brain (Corbetta and Shulman, 2002) distinguishes between more ventral stimulus-driven attentional systems and dorsal goal-directed control systems. In the context of affective processes, fast signatures of motivated attention can nicely be reconciled with the stimulus driven ventral system, while the dorsal structures, especially the dorsolateral prefrontal cortex have been linked to processes of top-down emotion regulation including attentional control over emotional distraction (Bishop et al., 2004). Support for a causal influence of dorsal structures on motivated attention comes for instance from studies that applied repetitive transcranial magnetic stimulation (rTMS) on dorsolateral prefrontal structures (Keuper et al., 2018; Notzon et al., 2018; Roesmann et al., 2019; Zwanzger et al., 2014). An inhibition of the dorsolateral prefrontal cortex was repeatedly shown to result in stronger EEG/MEG signatures of motivated attention in ventral brain regions at early and mid-latency, relatively automatic stages of stimulus processing (Keuper et al., 2018; Zwanzger et al., 2014). Considering the early onset of TF+/TF- discrimination in dorsal brain regions (43–83 ms), it seems possible that effects of generalized motivated attention in ventral brain regions (starting at 67–90 ms) were modulated by the inhibitory influence of 'associational' dorsal frontoparietal regions. However, a potential causal role of frontal functioning on the strength of generalized signatures of motivated attention still needs to be investigated, for example by means

of targeted non-invasive neurostimulation techniques. Overall, the observed interplay between dorsal and ventral brain networks fits well with the multiple-wave-model of emotion processing (Pessoa and Adolphs, 2010), which not only assumes multiple waves of activation within the visual stream, but also describes interactive connections between the visual cortex and 'associational' frontal and parietal regions.

5.4. Implications for existing models of fear generalization

At first sight, the early onset of neural generalization effects in ventral, visual regions fits well with the perceptual model of fear generalization (Lissek et al., 2014a), which conceptualizes fear generalization as a direct consequence of perceptual similarity or insufficient differentiation. This model claims that fear generalization arises from interactions in distributed networks, including regions of "fear--excitation" (including the amygdala) and prefrontal regions of "fear--inhibition". It also assigns a key role to the hippocampus and visual cortical areas, which are thought to perform comparisons between test stimuli (GS) and a memory template of the threat-signaling CS+. In case of increased perceptual similarity between GS and CS+ the result would be propagated to areas associated with both fear excitation and inhibition (Lissek et al., 2014a). Thus, the early effects in anterior temporal and visual areas fit well with the predicted role of the visual cortex in perceptual comparisons. However, the very early onset of "inhibitory" generalization effects in dorsal fronto-parietal networks (43–83 ms), and especially early discrimination effects in occipital regions (110–167 ms; 330–431 ms), revealing a clear neural differentiation of CS+ and all GS+ stimuli, challenge the key assumption of this model that fear generalization is a mere consequence of perceptual CS+/GS similarity.

In line with the observed selective CS+ processing in occipital regions, ssVEP and fMRI studies have previously shown that generalized behavioral, neural and autonomic measures may coexist with highly discriminative neural profiles (McTeague et al., 2015; Onat and Büchel, 2015). Particularly, CS+ specific enhancements of ssVEPs over central occipital sites (McTeague et al., 2015) fit nicely with the occipital effects reported here. Based on the finding that the integration of neural discrimination on the one hand and ambiguity-based uncertainty (Onat and Büchel, 2015) on the other hand predicted generalization effects in behavioral responses, Onat and Büchel (2015) conceptualized fear-generalization as an active, integrative process rather than a passive perception-driven one. This model might imply a temporal sequence of discrimination processes preceding generalization processes. However, our data rather suggest that the onset of generalization preceded the onset of discrimination. Possibly, the active integration of neural CS+/GS discrimination and uncertainty regarding the affective value of GS (Onat and Büchel, 2015) coexists with other mechanisms resulting in generalized fear responses.

To sum up, existing models cannot easily account for the temporal sequence of the different effects observed in the current study, which reveal the following key aspects: First, generalization is not (exclusively) based on the failure to discriminate. Second, processes of perceptual discrimination and generalization both affect very early stages of stimulus perception. Third, generalization effects may precede discrimination effects and might thus be (partly) independent from discriminatory perceptual functions. This last aspect is further supported by evidence suggesting that fear generalization in humans is not limited to perceptually similar stimuli (Dunsmoor and Murphy, 2015). Perception-independent processes such as category-based induction (Dunsmoor et al., 2014) and representations of conceptual knowledge (Dunsmoor and Murphy, 2014) were also shown to induce fear generalization. One possibility to account for early perceptual effects of generalization might be the assumption of plastic "mnemonic templates" representing the TF. Such templates might include perceptual, semantic and conceptual features, and might quickly be activated by potentially harmful new stimuli that are similar to threat-related stimuli (i.e. GS) without a full perceptual analysis. An integration of (non-perceptual) processes (e.g. "mnemonic templates") into existing perceptual models of fear generalization might be helpful to account for fast and implicit generalization effects.

5.5. General implications and future directions

Irrespectively of the underlying mechanisms, behavioral and neural generalization effects suggest that (1) newly acquired defensive responses to threat-signaling faces (CS+) quickly generalize to new identities (GS+), that (2) these generalization processes do not depend on awareness of the threat-related feature or dimension and that (3) they may be partly independent of the ability to discriminate. The balance of fear generalization on the one hand and the ability to discriminate harmful and safe stimuli on the other is regarded an adaptive mechanism which may facilitate survival-promoting reactions when facing potential danger. However, negative consequences of a dysfunctional interplay are manifold: First, overgeneralization of fear to harmless stimuli can be a burden to daily life and constitutes a key characteristic of anxiety and stress-related disorders (Dunsmoor and Paz, 2015). As previous clinical studies have mainly investigated overgeneralization of fear along obvious threat-related dimensions (Greenberg et al., 2013; Lissek et al., 2010), it remains to be disentangled whether pathological overgeneralization in anxiety disorders is grounded in alterations of more explicit or more implicit mechanisms. It seems likely that different groups of anxiety-related disorders might even depend on explicit and implicit generalization mechanisms to different degrees (Grillon et al., 2004). Future research should address these questions, because a better understanding of pathological mechanisms may help to improve therapeutic strategies aiming to reduce overgeneralization in anxiety disorders.

Second, our findings are in line with a previous study revealing that awareness of newly acquired CS/US contingencies is not necessary to modulate social biases to specific ethnic groups (Olson and Fazio, 2006). While Olson and Fazio (2006) showed that pre-existing biases can be *reduced* by means of implicit conditioning approaches, our study experimentally *induced* these biases. Both studies converge in the notion that even implicitly acquired associations of specific features with emotionally relevant events may result in generalized affective responses to newly encountered individuals sharing those specific features. Thus, implicit fear generalization and resulting biases affecting early stages of perception likely underpin the evolution of social prejudices and biases (c.f. De Houwer et al., 2001; Olson and Fazio, 2006). Notably, there is evidence that fear generalization and consequential social biases are modifiable. Discrimination trainings prior the assessment of fear generalization for instance may causally influence fear-related responses to generalization stimuli (e.g. Ginat-Frolich et al., 2019; Ginat-Frolich et al., 2017; Lommen et al., 2017). Moreover, if fearful responses rapidly generalize after aversive conditioning of implicit threat-relevant facial features, conditioned appetitive associations (Tapia León, Kruse, Stalder, Stark and Klucken, 2018) of these features might also be generalized or block fear acquisition and generalization. Thus, it would be interesting to study effects of explicit and implicit feature-based discrimination trainings as well as effects of explicit and implicit appetitive conditioning on generalized fear responses.

5.6. Limitations

Aiming at the investigation of implicit fear generalization, the confirmed TF+/US unawareness was the critical factor in this study. However, when evaluating the "degree of implicitness" one needs to acknowledge that participants in this study were rather aware of CS+/US associations of individual faces, as they correctly matched CS+ stimuli with US above chance level. Compared to previous visual MultiCS studies without CS+/US contingency awareness (Rehbein et al., 2014, 2015; Steinberg et al., 2012, 2013b), here the number of CS was smaller and CS/US associations were more frequent, because CS+ stimuli were

continuously reinforced in the conditioning and the generalization phase. Future studies with both absent contingency awareness of TF+/US *and* CS+/US associations are needed to support our hypothesis that non-conscious generalization does not depend on any explicit memory traces. Towards this aim, it is also important to carefully consider potential effects of task reactivity on (subsequent) behavioral and neurophysiological tests (cf. Lonsdorf et al., 2017). We acknowledge that the employed assessment modalities (e.g. offline assessment of behavioral tasks) and the order of tasks might have influenced the observed effects of contingency awareness and generalization. Nevertheless, under consideration of these limitations, our study demonstrates that MultiCS conditioning can facilitate the investigation of fear circuits that operate non-consciously. Importantly, the study of fear conditioning in absence of contingency awareness can support the translational transfer from animal models to models of human fear and anxiety. Such approach has been proposed as a way forward (LeDoux, 2014) to understand mechanisms of fear generalization in anxiety patients, who often show no awareness of fear evoking factors.

A second limitation of our study lies in the absence of a behavioral sensory discrimination task to directly assess discrimination performance regarding the employed stimuli. Generalization studies typically use slightly modified CS as generalization stimuli, e.g. a CS face or CS tone which was slightly morphed into a GS face or a GS tone (Laufer et al., 2016; Onat and Büchel, 2015). As aversive conditioning may reduce individual just noticeable differences between individual generalization stimuli that are very similar to the threat-signaling CS+ (Laufer et al., 2016; Shalev et al., 2018), studies employing such stimuli should test participants' ability to perceptually discriminate CS and GS *before* conditioning. Our study differs from these generalization experiments, as CS and GS stimuli were completely different faces, which just shared the abstract TF-low or TF-high. Therefore, a perceptual discrimination of individual faces within and across CS and GS sets could be assumed. Upon explicit instructions regarding the relevant facial feature, a perceptual discrimination of TF-high and TF-low faces would also be expected in discrimination tasks, because eye and mouth regions were clearly visible in all faces. However, in order to identify (and exclude) participants with specific discrimination disabilities or to associate pre-existing discrimination performance with measures of generalization, future studies in this direction should test participants' discrimination ability, for example in a behavioral forced-choice pair comparison task, comparing different faces regarding their TF. Overall, future Multi-CS conditioning studies should adopt sensory discrimination tasks to investigate the interplay of explicit and implicit aversive learning mechanisms, perceptual discrimination and fear generalization more directly.

6. Conclusion

Our results show that early amplifications of motivated attention towards conditioned faces in ventral brain regions generalize to novel individuals sharing a threat-related feature. This generalization process occurs in absence of TF/US-contingency awareness and may thus contribute to implicit attentional biases. Reverse effects in dorsal brain regions point towards early top-down modulations. The coexistence of generalized fear responses (indicated by similar CS+ and GS+ responses), and a clear discrimination (indicated by different CS+ and GS+ responses) in neurophysiological and behavioral data reveals it unlikely that the observed generalization effects reflect a mere failure of perceptual discrimination but rather supports an active, integrative process.

Declaration of competing interest

The authors declare no conflict of interest

Acknowledgments

This work was supported by the fund "Innovative Medical Research" of the University of Münster Medical School (RO211907) and the Collaborative Research Center (SFB-TRR58/C08)

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116302.

References

Andreatta, M., Leombruni, E., Glotzbach-Schoon, E., Pauli, P., Mühlberger, A., 2015. Generalization of contextual fear in humans. Behav. Ther. 46 (5), 583–596. https://doi.org/10.1016/j.beth.2014.12.008.

Bishop, S.J., Duncan, J., Brett, M., Lawrence, A.D., 2004. Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. Nat. Neurosci. 7 (2), 184–188. https://doi.org/10.1038/nn1173.

Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry 25 (1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9.

Bradley, M.M., Lang, P.J., 1999. International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings. Tech. Rep. No. B-2.

Bradley, M.M., Sabatinelli, D., Lang, P.J., Fitzsimmons, J.R., King, W., Desai, P., 2003. Activation of the visual cortex in motivated attention. Behav. Neurosci. 117 (2), 369–380. https://doi.org/10.1037/0735-7044.117.2.369. Retrieved from.

Bröckelmann, 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. J. Neurosci. 31 (21), 7801–7810. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/21613493.

Büchel, C., Dolan, R.J., 2000. Classical fear conditioning in functional neuroimaging. Curr. Opin. Neurobiol. 10 (2), 219–223. https://doi.org/10.1016/S0959-4388(00)00078-7.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3 (3), 215–229. https://doi.org/10.1038/nrn755.

De Houwer, J., Thomas, S., Baeyens, F., 2001. Association learning of likes and Dislikes : a review of 25 Years of research on human evaluative conditioning. Psychol. Bull. 127 (6), 853–869.

Dugas, M.J., Buhr, K., Ladouceur, C.D., 2004. In: Heimberg, R.G., Turk, C.L., Mennin, D.S. (Eds.), The Role of Intolerance of Uncertainty in Etiology and Maintenance. Guilford Press, New York.

Dunsmoor, J.E., Kragel, P.A., Martin, A., La Bar, K.S., 2014. Aversive learning modulates cortical representations of object categories. Cerebr. Cortex 24 (11), 2859–2872. https://doi.org/10.1093/cercor/bht138.

Dunsmoor, J.E., Murphy, G.L., 2014. Stimulus typicality determines how broadly fear is generalized. Psychol. Sci. 25 (9), 1816–1821. https://doi.org/10.1177/0956797614535401.

Dunsmoor, J.E., Murphy, G.L., 2015. February). Categories, concepts, and conditioning: how humans generalize fear. Trends In Cognitive Sciences. https://doi.org/10.1016/j.tics.2014.12.003. NIH Public Access.

Dunsmoor, J.E., Paz, R., 2015. Fear generalization and anxiety: behavioral and neural mechanisms. Biol. Psychiatry 78 (5), 336–343. https://doi.org/10.1016/j.biopsych.2015.04.010.

Dymond, S., Dunsmoor, J.E., Vervliet, B., Roche, B., 2015. Fear generalization in humans: systematic review and implications for anxiety disorder research. Behav. Ther. 46 (5), 561–582. https://doi.org/10.1016/j.beth.2014.10.001.

Gerlach, A.L., Andor, T., Patzelt, J., 2008. Die Bedeutung von Unsicherheitsintoleranz für die Generalisierte Angststörung Modellüberlegungen und Entwicklung einer deutschen Version der Unsicherheitsintoleranz-Skala. Z. Klin. Psychol. Psychother. 37 (3), 190–199. https://doi.org/10.1026/1616-3443.37.3.190.

Ginat-Frolich, R., Gendler, T., Marzan, D., Tsuk, Y., Shechner, T., 2019. Reducing fear overgeneralization in children using a novel perceptual discrimination task. Behav. Res. Ther. 116, 131–139. https://doi.org/10.1016/j.brat.2019.03.008.

Ginat-Frolich, R., Klein, Z., Katz, O., Shechner, T., 2017. A novel perceptual discrimination training task: reducing fear overgeneralization in the context of fear learning. Behav. Res. Ther. 93 (March), 29–37. https://doi.org/10.1016/j.brat.2017.03.010.

Greenberg, T., Carlson, J.M., Cha, J., Hajcak, G., Mujica-Parodi, L.R., 2013. Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. Depress. Anxiety 30 (3), 242–250. https://doi.org/10.1002/da.22016.

Grillon, C., Baas, J.P., Lissek, S., Smith, K., Milstein, J., 2004. Anxious responses to predictable and unpredictable aversive events. Behav. Neurosci. 118 (5), 916–924. https://doi.org/10.1037/0735-7044.118.5.916.

Haddad, A.D.M., Pritchett, D., Lissek, S., Lau, J.Y.F., 2012. Trait anxiety and fear responses to safety cues: stimulus generalization or sensitization? J. Psychopathol. Behav. Assess. 34, 323–331. https://doi.org/10.1007/s10862-012-9284-7.

Haddad, A.D.M., Xu, M., Raeder, S., Lau, J.Y.F., 2013. Measuring the role of conditioning and stimulus generalisation in common fears and worries. Cognit. Emot. 27 (5), 914–922. https://doi.org/10.1080/02699931.2012.747428.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32 (1), 35–42. Retrieved from. https://www.springerlink.com/index/4474345W652H8581.pdf.

Hauk, O., 2004. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. Neuroimage 21 (4), 1612–1621. https://doi.org/10.1016/J.NEUROIMAGE.2003.12.018.

Hautzinger, M., Bailer, M., 1993. Allgemeine Depressoins Skala - ADS. Beltz, Weinheim.

Hintze, P., Junghöfer, M., Bruchmann, M., 2014. Evidence for rapid prefrontal emotional evaluation from visual evoked responses to conditioned gratings. Biol. Psychol. 99, 125–136. Retrieved from. https://www.elsevier.com/authorsrights.

Junghofer, M., Bradley, M.M., Elbert, T.R., Lang, P.J., 2001. Fleeting images: a new look at early emotion discrimination. Psychophysiology 38 (2), 175–178. https://doi.org/10.1111/1469-8986.3820175.

Junghöfer, M., Elbert, T., Tucker, D.M., Rockstroh, B., 2000. Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology 37 (4), 523–532. https://doi.org/10.1111/1469-8986.3740523.

Junghöfer, M., Rehbein, M.A., Maitzen, J., Schindler, S., Kissler, J., 2016. An evil face? -Verbal evaluative Multi-CS conditioning enhances face-evoked mid- latency magnetoencephalographic responses. Soc. Cogn. Affect. Neurosci. 12 (4), 695–705. Retrieved from. https://scan.oxfordjournals.org/.

Keuper, K., Terrighena, E., Chan, C.C.H., Junghöfer, M., Lee, T.M.C., 2018. How the dorsolateral prefrontal cortex controls affective processing in absence of visual awareness – insights from a combined EEG-rTMS study. Front. Hum. Neurosci. 12, 412. https://doi.org/10.3389/fpls.2018.01222.

Keuper, K., Zwanzger, P., Nordt, M., Eden, A., Laeger, I., Zwitserlood, P., et al., 2014. How "love" and "hate" differ from "sleep": using combined electro/magnetoencephalographic data to reveal the sources of early cortical responses to emotional words. Hum. Brain Mapp. 35 (3), 875–888. https://doi.org/10.1002/hbm.22220.

Kim, J.J., Jung, M.W., 2006. Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. Neurosci.BioBehav. Rev.NIH.Public Access 30 (2), 188–202. https://doi.org/10.1016/j.neubiorev.2005.06.005.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A., 2010. Presentation and validation of the radboud faces database. Cognit. Emot. 24 (8), 1377–1388. https://doi.org/10.1080/02699930903485076.

Laufer, O., Israeli, D., Paz, R., 2016. Behavioral and neural mechanisms of overgeneralization in anxiety. Curr. Biol. 26 (6), 713–722. https://doi.org/10.1016/j.cub.2016.01.023.

Laux, L., Glanzmann, P., Schaffner, P., Spielberger, C.D., 1981. Das State-Trait-Angstinventar (STAI): Theoretische Grundlagen und Handanweisung. Beltz, Weinheim. Retrieved from. https://opus4.kobv.de/opus4-bamberg/frontdoor/index/index/docId/31823.

LeDoux, J.E., 2014. Coming to terms with fear. Proc. Natl. Acad. Sci. 111 (8) https://doi.org/10.1073/pnas.1400335111.

Lissek, S., Biggs, A.L., Rabin, S.J., Cornwell, B.R., Alvarez, R.P., Pine, D.S., Grillon, C., 2008. Generalization of conditioned fear-potentiated startle in humans: experimental validation and clinical relevance. Behav. Res. Ther. 46 (5), 678–687.

Lissek, S., Bradford, D.E., Alvarez, R.P., Burton, P., Espensen-Sturges, T., Reynolds, R.C., Grillon, C., 2014a. Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. Soc. Cogn. Affect. Neurosci. 9 (8), 1134–1142. https://doi.org/10.1093/scan/nst096.

Lissek, S., Kaczkurkin, A.N., Rabin, S., Geraci, M., Pine, D.S., Grillon, C., 2014b. Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. Biol. Psychiatry 75 (11), 909–915. https://doi.org/10.1016/j.biopsych.2013.07.025.

Lissek, S., Rabin, S., Randi, H.E., Lukenbaugh, D., Geraci, M., Pine, D.S., Grillon, C., 2010. Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. Am. J. Psychiatry 167 (1), 47–55. Retrieved from. https://clinicaltrials.gov/show/NCT02148042.

Lommen, M.J.J., Duta, M., Vanbrabant, K., De Jong, R., Juechems, K., Ehlers, A., 2017. Training discrimination diminishes maladaptive avoidance of innocuous stimuli in a fear conditioning paradigm. PLoS One 12 (10), 1–14. https://doi.org/10.1371/journal.pone.0184485.

Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., et al., 2017. Don't fear 'fear conditioning': methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci. Biobehav. Rev. 77, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026.

Lundqvist, D., Flykt, A., Öhman, A., 1998. The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience. Psychology section. Karolinska Institutet 91, 630.

MacNamara, A., Hajcak, G., 2009. Anxiety and spatial attention moderate the electrocortical response to aversive pictures. Neuropsychologia 47 (13), 2975–2980. https://doi.org/10.1016/j.neuropsychologia.2009.06.026.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

McTeague, L.M., Gruss, L.F., Keil, A., 2015. Aversive learning shapes neuronal orientation tuning in human visual cortex. Nat. Commun. 6, 7823. https://doi.org/10.1038/ncomms8823.

Miskovic, V., Keil, A., 2012. Acquired fears reflected in cortical sensory processing: a review of electrophysiological studies of human classical conditioning. Psychophysiology 49 (9), 1230–1241. https://doi.org/10.1111/j.1469-8986.2012.01398.x.

Nelson, B.D., Weinberg, A., Pawluk, J., Gawlowska, M., Proudfit, G.H., 2015. An event-related potential investigation of fear generalization and intolerance of uncertainty. Behav. Ther. 46 (5), 661–670. https://doi.org/10.1016/j.beth.2014.09.010.

Notzon, S., Steinberg, C., Zwanzger, P., Junghöfer, M., 2018. Modulating emotion perception – opposing effects of inhibitory and excitatory prefrontal cortex stimulation. Biol. Psychiatry: Cogn. Neurosci.Neuroimaging 3 (4), 329–336. https://doi.org/10.1016/j.bpsc.2017.12.007.

Öhman, A., Mineka, S., 2001. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychol. Rev. 108 (3), 483–522. https://doi.org/10.1037//0033-295X.108.3.483.

Olofsson, J., Polich, J., 2007. Affective visual event-related potentials: arousal, repetition, and time-on-task. Biol. Psychol. 75 (1), 101–108. Retrieved from. https://www.sciencedirect.com/science/article/pii/S0301051107000026.

Olson, M.A., Fazio, R.H., 2006. Reducing automatically activated racial prejudice through implicit evaluative conditioning. Personal. Soc. Psychol. Bull. 32 (4), 421–433. https://doi.org/10.1177/0146167205284004.

Onat, S., Büchel, C., 2015. The neuronal basis of fear generalization in humans. Nat. Neurosci. 18 (12), 1811–1818. https://doi.org/10.1038/nn.4166.

Pastor, M.C., Rehbein, M.A., Junghöfer, M., Poy, R., LÃpez, R., MoltÃ, J., 2015. Facing challenges in differential classical conditioning research: benefits of a hybrid design for simultaneous electrodermal and electroencephalographic recording. Front. Hum. Neurosci. 9 (June), 1–14. https://doi.org/10.3389/fnhum.2015.00336.

Pavlov, I.P., 1927. Conditional Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex. Oxford Univ. Press, Oxford.

Pessoa, L., Adolphs, R., 2010. Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. Nat. Rev. Neurosci. 11 (11), 773–783. https://doi.org/10.1038/nrn2920.

Peyk, P., De Cesarei, A., Junghöfer, M., 2011. ElectroMagnetoEncephalography software: overview and integration with other EEG/MEG toolboxes. Comput. Intell. Neurosci. 2011, 861705. https://doi.org/10.1155/2011/861705.

Rehbein, M.A., Pastor, M.C., Moltó, J., Poy, R., López-Penadés, R., Junghöfer, M., 2018. Identity and expression processing during classical conditioning with faces. Psychophysiology 55 (10). https://doi.org/10.1111/psyp.13203.

Rehbein, M.A., Steinberg, C., Wessing, I., Pastor, M.C., Zwitserlood, P., Keuper, K., Junghöfer, M., 2014. Rapid plasticity in the prefrontal cortex during affective associative learning. PLoS One 9 (10). https://doi.org/10.1371/journal.pone.0110720.

Rehbein, M.A., Wessing, I., Zwitserlood, P., Steinberg, C., Eden, A.S., Dobel, C., Junghöfer, M., 2015. Rapid prefrontal cortex activation towards aversively paired faces and enhanced contingency detection are observed in highly trait-anxious women under challenging conditions. Front. Behav. Neurosci. 9 (June), 1–19. https://doi.org/10.3389/fnbeh.2015.00155.

Roesmann, K., Dellert, T., Junghoefer, M., Kissler, J., Zwitserlood, P., Zwanzger, P., Dobel, C., 2019. The causal role of prefrontal hemispheric asymmetry in valence processing of words – insights from a combined cTBS-MEG study. Neuroimage 191, 367–379. https://doi.org/10.1016/j.neuroimage.2019.01.057.

Sabatinelli, D., Lang, P.J., Keil, A., Bradley, M.M., 2007. Emotional perception: correlation of functional MRI and event-related potentials. Cerebr. Cortex 17 (5), 1085–1091. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/16769742.

Schupp, H.T., Cuthbert, B., Bradley, M.M., Hillman, C., Hamm, A., Lang, P.J., 2004. Brain processes in emotional perception: motivated attention. Cognit. Emot. 18 (5), 593–611. Retrieved from. https://www.tandfonline.com/doi/abs/10.1080/02699930341000239.

Schupp, H.T., Flaisch, T., Stockburger, J., Junghöfer, M., 2006. Emotion and attention: event-related brain potential studies. Progress in brain research 156, 31–51.

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., Konrad, C., 2009. June 10). In: Gendelman, H.E. (Ed.), Human Fear Conditioning and Extinction in Neuroimaging: A Systematic Review, PLoS ONE. Public Library of Science.

Shalev, L., Paz, R., Avidan, G., 2018. Visual aversive learning compromises sensory discrimination. J. Neurosci.: Off. J.Soc. Neurosci 38 (11), 2766–2779. https://doi.org/10.1523/JNEUROSCI.0889-17.2017.

Steinberg, C., Bröckelmann, A.K., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., Junghöfer, M., 2013a. Preferential responses to extinguished face stimuli are preserved in frontal and occipito-temporal cortex at initial but not later stages of processing. Psychophysiology 50 (3), 230–239. https://doi.org/10.1111/psyp.12005.

Steinberg, C., Bröckelmann, A.K., Rehbein, M.A., Dobel, C., Junghöfer, M., 2013b. Rapid and highly resolving associative affective learning: convergent electro- and magnetoencephalographic evidence from vision and audition. Biol. Psychol. 92 (3), 526–540. Retrieved from. https://linkinghub.elsevier.com/retrieve/pii/S0301051112000312.

Steinberg, C., Dobel, C., Schupp, H.T., Kissler, J., Elling, L., Pantev, C., Junghöfer, M., 2012. Rapid and highly resolving: affective evaluation of olfactorily conditioned faces. J. Cogn. Neurosci. 24 (1), 17–27. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/21671742.

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. Cerebr. Cortex 16 (6), 876–887. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/16151178.

Tabbert, K., Merz, C.J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O.T., Stark, R., 2011. Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. Soc. Cogn. Affect. Neurosci. 6 (4), 495–506. https://doi.org/10.1093/scan/nsq070.

Tabbert, K., Stark, R., Kirsch, P., Vaitl, D., 2006. Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. Neuroimage 32 (2), 761–770. https://doi.org/10.1016/j.neuroimage.2006.03.038.

Tapia León, I., Kruse, O., Stalder, T., Stark, R., Klucken, T., 2018. Neural correlates of subjective CS/UCS association in appetitive conditioning. Hum. Brain Mapp. 39 (4), 1637–1646. https://doi.org/10.1002/hbm.23940.

Vervliet, B., Geens, M., 2014. Fear generalization in humans: impact of feature learning on conditioning and extinction. Neurobiol. Learn. Mem. 113, 143–148. https://doi.org/10.1016/j.nlm.2013.10.002.

Vervliet, B., Kindt, M., Vansteenwegen, D., Hermans, D., 2010. Fear generalization in humans: impact of verbal instructions. Behav. Res. Ther. 48 (1), 38–43. https://doi.org/10.1016/j.brat.2009.09.005.

Vuilleumier, P., 2005. How brains beware: neural mechanisms of emotional attention. Trends Cogn. Sci. 9 (12), 585–594. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/16289871.

Wickens, T.D., 2002. Elementary Signal Detection Theory. Oxford University Press, USA.

Zwanzger, P., Steinberg, C., Rehbein, M.A., Bröckelmann, A.K., Dobel, C., Zavorotnyy, M., Junghöfer, M., 2014. Inhibitory repetitive transcranial magnetic stimulation (rTMS) of the dorsolateral prefrontal cortex modulates early affective processing. Neuroimage 101, 193–203. https://doi.org/10.1016/j.neuroimage.2014.07.003.