

## REGULAR RESEARCH ARTICLE

# Stimulus-Based Extinction Generalization: Neural Correlates and Modulation by Cortisol

Bianca Hagedorn, Oliver T. Wolf, Christian J. Merz<sup>o</sup>

Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University, Bochum, Germany.

Correspondence: Dr. Christian J. Merz, Department of Cognitive Psychology, Faculty of Psychology, Institute of Cognitive Neuroscience, Ruhr University Bochum, Universitätsstr. 150, 44780 Bochum, Germany ([Christian.J.Merz@rub.de](mailto:Christian.J.Merz@rub.de)).

## Abstract

**Background:** While healthy individuals and patients with anxiety disorders easily generalize fear responses, extinction learning is more stimulus specific. Treatments aiming to generalize extinction learning are urgently needed, since they comprise the potential to overcome stimulus specificity and reduce relapses, particularly in the face of stressful events.

**Methods:** In the current 3-day functional magnetic resonance imaging fear conditioning paradigm, we aimed to create a generalized extinction memory trace in 60 healthy men and women by presenting multiple sizes of 1 conditioned stimulus during extinction training (CS+G; generalized), whereas the other conditioned stimulus was solely presented in its original size (CS+N; nongeneralized). Recall was tested on the third day after pharmacological administration of either the stress hormone cortisol or placebo.

**Results:** After successful fear acquisition, prolonged activation of the amygdala and insula and deactivation of the ventromedial prefrontal cortex for CS+G compared with CS+N during extinction learning indicated sustained fear to the generalization stimuli. In line with our hypotheses, reduced amygdala activation was observed after extinction generalization on the third day in the contrast CS+G minus CS+N, possibly reflecting an attenuated return of fear. Cortisol administration before recall, however, blocked this effect.

**Conclusions:** Taken together, the findings show that extinction generalization was associated with decreased activation of the fear network during recall after prolonged activation of the fear network during extinction learning. However, the generalization of the extinction memory did not counteract the detrimental effects of stress hormones on recall. Thus, stimulus-based extinction generalization may not be sufficient to reduce relapses after stressful experiences.

**Key Words:** Fear conditioning, functional magnetic resonance imaging, glucocorticoids, return of fear, stress hormones

## Introduction

Although exposure therapy is commonly regarded as the first-line treatment for anxiety disorders, patients suffer from relapses after successful therapy (Craske et al., 2006), especially after stressful events (Jacobs and Nadel, 1985; Francis et al., 2012; de Quervain et al., 2019). Extinction learning represents the major underlying mechanism for exposure therapy; thus, principles strengthening extinction learning should also enhance exposure therapy (Scheveneels et al., 2016; Forcadell et al., 2017;

Craske et al., 2018; Lange et al., 2020). We will focus on stimulus generalization during extinction training as a promising candidate to increase extinction learning and to prevent the return of fear.

In anxiety disorders, fear is not only evoked by the original fear-related stimulus but also by a multitude of perceptually (Holt et al., 2014; Struyf et al., 2015, 2017; Zaman et al., 2019) and conceptually (Dunsmoor et al., 2011, 2012; Vervoort et al.,

Received: April 8, 2020; Revised: October 7, 2020; Accepted: November 10, 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of CINP.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Significance Statement

This study aimed to identify the neural correlates of stimulus-based extinction generalization for the first time, to our knowledge, which is one promising approach to be incorporated in exposure therapy to reduce relapses. To investigate the effects that are solely attributable to extinction generalization, classical extinction training with the original conditioned stimulus only was compared with extinction training with multiple generalization stimuli. Extinction training with multiple conditioned stimuli in addition to the original conditioned stimulus led to prolonged activation of the amygdala and insula and deactivation of the vmPFC. However, extinction generalization caused reduced activation of the fear network during recall on the following day. Interestingly, the administration of the stress hormone cortisol before recall seemed to block the effects of extinction generalization, which can thus serve as a potential model for stress-induced relapses.

2014; Bennett et al., 2015; Dunsmoor and Murphy, 2015; Starita et al., 2019) related stimuli. Therefore, fear seems to generalize naturally over a variety of related stimuli (Lissek et al., 2014; de Voogd et al., 2020), whereas extinction acts more specifically (Bouton, 2002, 2004; Vervliet et al., 2006; Bouton and Todd, 2014; Pappens et al., 2015). Fear generalization is assumed to rely on pattern separation and pattern completion mechanisms in the hippocampus (Lissek et al., 2014; Asok et al., 2019) matching the encountered stimulus to the original feared stimulus. If the schematic matching process initiates pattern separation (discrimination between the stimuli), the ventromedial prefrontal cortex (vmPFC) will inhibit the fear network consisting of the amygdala, insula, and the dorsal anterior cingulate cortex (dACC). In contrast, pattern completion (missing discrimination between the stimuli) should result in a reduced inhibitory control of the vmPFC that in turn leads to enhanced activation of the fear network (Lissek et al., 2014). However, it remains elusive whether extinction generalization follows the same principles as fear generalization.

In contrast to extinction training including the original conditioned stimulus (CS) only, using multiple stimuli or contexts during extinction training might create a more generalized extinction memory trace. Nevertheless, incorporating multiple stimuli or contexts omitting the original CS during extinction training resulted in less successful extinction learning for the original CS (Vervliet et al., 2006; Xu et al., 2018; Wong and Lovibond, 2020). In contrast, the utilization of multiple stimuli encompassing the original CS for exposure therapy in a spider-fearful sample resulted in prolonged fear expression during exposure therapy but eventually led to less return of fear (Rowe and Craske, 1998). Although context generalization is assumed to constitute an important factor for the long-term effect of exposure therapy, stimulus-based generalization stimuli have also been successfully incorporated in treatment analog studies and represent an important part of exposure therapy (Craske et al., 2014). Interestingly, the most beneficial effects in phobic patients, as shown in reduced short- and long-term return of fear, were observed when multiple stimuli were presented during extinction training in contrast to multiple contexts or multiple stimuli and contexts (Shiban et al., 2015).

In the present study, new evidence and neural underpinnings of extinction generalization will be elucidated. As a first hypothesis, enhanced activation of the vmPFC is expected to downregulate activation of the fear network during recall of the generalized extinguished conditioned stimulus (CS+G) relative to a classically extinguished, nongeneralized conditioned stimulus (CS+N), mirrored in decreased skin conductance responses (SCRs) to the CS+G compared with CS+N. These effects are expected to rely on enhanced attentional processes during extinction training towards the CS+G compared with the CS+N due to heightened arousal or novelty. The hippocampus

is expected to mediate these effects via pattern completion; therefore, enhanced activation of the hippocampus during recall of the generalized compared with the nongeneralized extinguished conditioned stimulus is assumed. Correspondingly, the increased variability during extinction training could enhance the prediction error occurring for the CS+G compared with the CS+N. Exploratory analyses will be conducted to investigate the influence of sex on extinction generalization (Merz et al., 2018b; Velasco et al., 2019).

Additionally, stress and the stress hormone cortisol have been shown to promote relapses after successful exposure therapy (Jacobs and Nadel, 1985; Francis et al., 2012) and likewise the return of fear after successful extinction learning (Deschaux et al., 2013; Hamacher-Dang et al., 2013; Raio et al., 2014; Meir Drexler et al., 2019). Furthermore, stress hormones promote the return of fear by increasing activation of the fear network (Kinner et al., 2018) while decreasing activation of the inhibitory extinction network, including the vmPFC (Kinner et al., 2016; Meir Drexler et al., 2019). As a second hypothesis, we will investigate whether extinction generalization is sufficiently robust to withstand the detrimental effects of stress hormones on the extinction memory trace, reflected in lower activation of the fear network and lower SCRs for CS+G compared with CS+N in the cortisol group due to increased responding to the CS+N. In addition, effects of sex on cortisol effects will be explored (Merz and Wolf, 2017).

## Methods

### Participants

According to the performed power analysis using G\*Power 3.1 (Faul et al., 2007) with an assumed small effect size of  $f=0.105$  determined by the effect of stress on recall processes (Shields et al., 2017), an assumed correlation of 0.8 between repeated measures and a significance level  $P=.05$ , 60 participants would be required to achieve a power of  $1-\beta=.80$ . Sixty healthy participants (30 women; mean age: 24.5 years, SD: 3.8, range: 18–35 years; mean BMI: 23.3 kg/m<sup>2</sup>, SD: 2.7, range: 18–28 kg/m<sup>2</sup>) recruited at the Ruhr University Bochum completed the experiment. Of 62 participants in total, 2 had to be excluded due to technical issues and missing contingency awareness. Women were not tested during pregnancy or menstruation, and women taking oral contraceptives were excluded to reduce alterations in circulating sex hormones and their impact on fear-conditioning processes (Merz et al., 2018b; Velasco et al., 2019). Controlling for menstrual cycle phase appeared not to be feasible due to practical reasons in this 3-day study, although endogenous estrogens are assumed to exert effects on fear-conditioning processes, especially extinction learning (Hwang et al., 2015; Merz et al., 2018b; Hammoud et al., 2020).

Participants reported normal or corrected-to-normal vision. In addition to the application of standard functional magnetic resonance imaging (fMRI) exclusion criteria, students with chronic or acute psychiatric or neurological illnesses, mental disorders, and regular or acute intake of medication were excluded. The Edinburgh Inventory of Handedness (Oldfield, 1971) confirmed right-handedness of all participants.

Individual appointments were scheduled between 1 PM and 8 PM to reduce circadian variations in endogenous cortisol levels (Chung et al., 2011). Participants were asked to refrain from exercising, drinking anything except water, and eating 2 hours prior to each session. All participants provided informed consent before the start of the experiment, were reimbursed with 45€, and were debriefed on the last day of the experiment. All procedures were approved by the local ethics committee of the medical faculty (registration no. 16–5789) and conducted in accordance with the Declaration of Helsinki.

### Fear-Conditioning Procedure

In the present differential fear-conditioning paradigm taking place on 3 consecutive days, 3 white geometrical shapes (square, rhomb, and parallelogram) with identical luminescence on a black background served as CS (see Figure 1). In each trial, the CS was shown for 8 seconds followed by a jittered 9.5- to 12-second black screen inter-trial interval. The black screen presented at the beginning of each trial was jittered between 0 and 2.5 seconds, resulting in a fixed trial duration of 20 seconds. The assignment of the shapes to the 3 CS was pseudo-randomized and balanced between groups. A 100-ms electrical stimulation served as UCS and was applied via two 1-cm<sup>2</sup> electrodes attached to the fingertips of the participant's right index and middle-finger using a constant voltage stimulator (STM200; BIOPAC Systems, CA). Stimulation level was individually set to be unpleasant but not painful (see Kinner et al., 2018; Merz et al., 2018a).

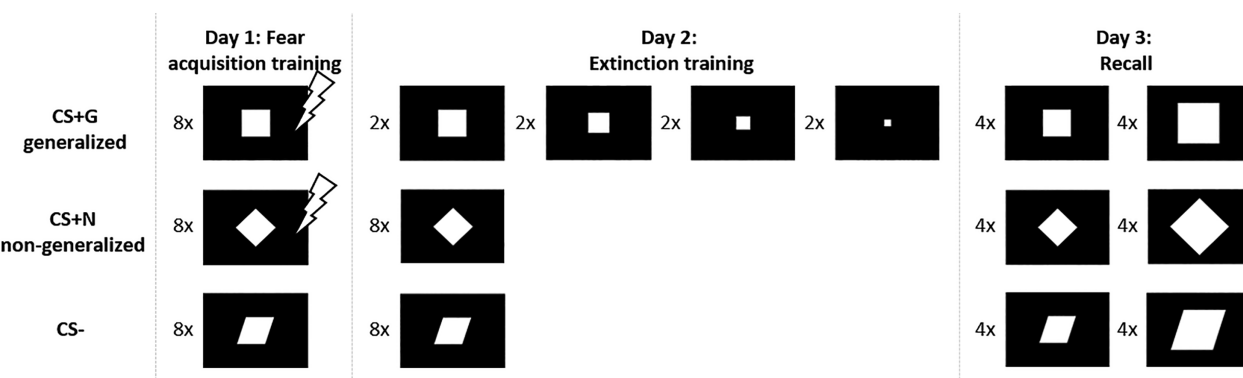
Prior to fear acquisition training, participants were instructed to pay attention to possible associations between the presentation of a geometrical shape and the electrical stimulation, as they would be asked about their observations afterwards. All 60 participants were classified as contingency aware, since they correctly indicated electrical stimulations following both CS+ and never following the CS– after fear acquisition

training (Tabbert et al., 2011). Prior to extinction training and recall, participants were informed that the acquired associations on day 1 would persist and not change during the experiment to avoid expectancy of contingency reversal. The participants were not informed about the actual contingencies over the course of the entire experiment.

During fear acquisition training on day 1, 8 trials were presented for each of the 3 CS; the 2 CS+ were immediately followed by the UCS with a partial reinforcement rate of 62.5% (5 of 8 trials), whereas the CS– was never reinforced (Figure 1). During extinction training on day 2, the CS– and the CS+N were presented solely in their original size, whereas the CS+G was presented in 3 smaller sizes (75%, 50%, and 25% of the original size) in addition to its original size. Each CS underwent 8 extinction trials (each of the 4 CS+G sizes was presented 2 times) without reinforcement. During recall on day 3, all stimuli were presented in 1 greater size (175%) for 4 trials intermixed with 4 presentations in their original size (see Figure 1). The greater size was implemented to test for pure generalization effects to a version of the CS+G not previously presented. Importantly, we decided to include the same number of presentations for each CS during extinction training and recall irrespective of size to avoid learning effects solely due to a higher number of presentations. After recall, 4 electrical stimulations were applied during reinstatement followed by a reinstatement test, which is not reported in the main manuscript (see supplementary information and supplementary Table 2). The paradigm was realized in Matlab 2017a (Mathworks Inc., Sherborn, MA), and stimuli were presented using MR-suitable LCD goggles (Visuastim Digital, Resonance Technology Inc., Northridge, CA).

### Cortisol Administration and Saliva Samples

In a randomized double-blind design, one half of the participants (15 women and 15 men) received two 10-mg hydrocortisone tablets (Hoechst) 40 minutes before recall on day 3, whereas the other half of the participants (15 women and 15 men) received visually identical placebos. Saliva samples were taken on day 3 using Salivettes (Sarstedt, Nümbrecht, Germany) prior to tablet intake as a baseline, 30 minutes, and 60 minutes after tablet intake (before and after recall). All saliva samples were stored at –20°C until analyzed with a commercial enzyme-linked immunosorbent assay (IBL



**Figure 1.** Extinction generalization paradigm. The 3 phases (fear acquisition training, extinction training, and recall) were performed on 3 consecutive days with less than a 2-hour shift. Lightning bolts represent electrical stimulation as unconditioned stimulus in reinforced conditioned stimulus (CS+) trials (62.5% partial reinforcement) during fear acquisition training. In total, there were 8 presentations of each of the 3 CS on each day, which were adjusted according to the number of versions for each geometrical shape. Thus, during extinction training, each of the 4 versions of the CS+G (generalized) was only presented 2 times, whereas the CS+N (nongeneralized) and the non-reinforced conditioned stimulus (CS–) was presented 8 times each. Accordingly, each of the 2 versions of each CS were presented 4 times during recall. The CS distribution was equal over the course of each phase with the restriction of no more than 2 consecutive trials of the same CS.

International, Hamburg, Germany). Inter- and intra-assay variations were <10%. One participant had to be excluded from the cortisol analysis only due to nondetectable low cortisol levels.

### Physiological Data

SCRs were measured with Ag/AgCl electrodes filled with an isotonic (0.05 NaCl) electrolyte medium attached to the hypothear of the left hand. Data were acquired at 5000 Hz using the Brain Vision Recorder software, filtered at 4.5 Hz, and resampled at 10 Hz in the Brain Vision Analyzer (Brain Products GmbH, Munich, Germany). Conditioned responses were analyzed via Ledalab 3.4.9 (Benedek and Kaernbach, 2010a, 2010b) and specified as trough-to-peak maximum amplitudes in a time window of 1 to 8 seconds after CS onset. Analyses were carried out with transformed values (natural logarithm) to attain normal distribution. Four participants were excluded from the SCR analysis only due to technical failure of the recording system at least on 1 day.

### Statistical Analyses

Statistical analyses of cortisol concentrations and SCRs were tested with mixed ANOVAs performed in IBM SPSS Statistics 21 (IBM Corp., Armonk, NY) with the significance threshold set to .05 (Bonferroni-corrected for multiple comparisons); effect sizes were reported in partial eta square ( $\eta_p^2$ ). Treatment (cortisol vs placebo) and sex (men vs women) and their interaction were always entered as between-participants variables. The within-participants factor time (baseline, 30 minutes vs 60 minutes after tablet intake) was entered for the analysis of cortisol concentrations. SCR analyses encompassed the comparison of the mean response to CS+G and CS+N against the CS- for fear acquisition training. For extinction training, each CS was entered separately, and the within-subjects factor half (first vs second half) was added. For recall, the within-participant factors CS and size (original vs modified) were included. Greenhouse-Geisser corrected values were reported if the assumption of sphericity was violated.

### Functional Magnetic Resonance Imaging

Whole-brain images were measured using a 3T whole-body scanner with a 32-channel head coil (Philips Achieva 3.0 T X-Series, Philips, the Netherlands). Structural images encompassed 220 transversally oriented slices (FOV: 240 mm × 240 mm, voxel size: 1 mm × 1 mm × 1 mm) obtained in a T1 weighted FTE sequence. Functional images encompassed 40 ascending slices measured parallel to the orbitofrontal bone transition (FOV: 192 mm × 192 mm, voxel size: 2 mm × 2 mm × 3 mm) obtained with a T2 weighted gradient echoplanar imaging sequence (TR: 2.5 seconds, TE: 30 milliseconds, flip angle: 67°, slice gap: 0.75 mm). During each scan session, 201 volumes were recorded, while 210 additional volumes were recorded for reinstatement and reinstatement test. In addition to the 3 dummy scans preceding each functional scan session, the first 3 functional images were discarded to reach stable magnetization.

The software Statistical Parametric Mapping (SPM12, Wellcome Department of Cognitive Neurology, London, UK) applied in Matlab 2017a (Mathworks Inc., Sherborn, MA) served for preprocessing and analyses of imaging data. Preprocessing contained realignment, slice time correction, co-registration to the participant's structural image, normalization to MNI standard space, and smoothing using an 8-mm FWHM Gaussian kernel.

In the first level model, the 3 scan sessions were entered separately for each participant. Regressors in each model encompassed CS types (CS+G, CS+N, and CS-). For each scan session, additional parameters were entered: fear acquisition parameters encompassed blocks (first and second half) as well as UCS transmission and omission (separately for each CS), while extinction parameters encompassed blocks (first and second half). Recall parameters included size (original and modified) as well as the UCS and the gray screen shown during reinstatement (see supplementary information). In addition, the 6 realignment parameters were included as covariates, and a high pass filter with a time constant of 128 seconds was applied. In the general linear model, all parameters were modeled using a stick function and convolved with the hemodynamic response function in an event-related design.

In the second level, full-factorial models with the factors treatment and sex were conducted. In line with SCR analyses, the contrast CS+G AND CS+N minus CS- served to examine successful fear learning due to missing differences between CS+G and CS+N within fear learning. To capture time-dependent changes during extinction learning between generalized and nongeneralized CS, the critical contrast CS+G minus CS+N was tested in the first and second block. The effectiveness of the extinction generalization procedure was analyzed with the contrast CS+G minus CS+N (both sizes) for recall. To compare the classical extinction protocol with the extinction protocol with multiple stimuli, the contrast CS+G minus CS+N is the most critical and direct test of our hypotheses. This approach also circumvents the possible problem of the usual comparison with the CS-, which is also learned as a safety signal (Lissek et al., 2005) and might obscure the result pattern.

As a first hypothesis, successful extinction generalization during recall should be reflected by less activation of fear-related areas (insula, amygdala, and dACC) and reduced SCRs towards both sizes of the CS+G compared with both sizes of the CS+N. This definition of generalization relies on the assumption of extinction generalization leading to enhanced extinction learning that is expected to decrease responding to the original as well as to the altered CS+G. Pure generalization effects, however, should be captured in the comparison of the modified size of the CS+G with the modified size of the CS+N during recall. Higher novelty during generalized extinction training is expected to reflect an increased prediction error tested in the contrast CS+G minus CS+N that enhances extinction recall. This effect should be indicated by higher activation of the fear network and lower activation of fear-inhibitory areas as well as increased SCRs toward CS+G vs CS+N during extinction training.

As a second hypothesis, cortisol effects will also be investigated within the contrast CS+G minus CS+N. We expect lower activation of the fear network and lower SCRs for CS+G compared with CS+N in the cortisol relative to the placebo group due to increased responding to the CS+N.

Region of interest (ROI) analyses encompassed regions identified and expected to be involved in fear generalization (Lissek et al., 2014) due to the assumption of comparable underlying neural mechanisms: insula, amygdala, dACC (fear excitation; Fullana et al., 2016), vmPFC (fear inhibition; Fullana et al., 2018), and hippocampus (mediating area via pattern separation/pattern completion; Lissek et al., 2014). Maximum probability masks (1 mm) from Harvard-Oxford Cortical- and Subcortical-Atlases with the threshold set to 0.25 were used for the insula, amygdala, and hippocampus. The dACC and vmPFC masks consisted of a 5-mm sphere around the peak voxel previously identified in meta-analyses regarding fear acquisition for dACC (MNI: x=0, y=16, z=36; Mechias et al., 2010) and regarding extinction for vmPFC



(MNI:  $x=0, y=40, z=-3$ ; Schiller and Delgado, 2010). Family-wise error (FWE) correction (Penny et al., 2007) for small volumes was applied to the significance threshold of  $P \leq .05$  for the predefined ROIs. The significance threshold was set to  $P \leq .05$  with family-wise error correction applied and a minimal cluster size of 10 voxels for exploratory whole-brain analyses. In addition, functional connectivity was investigated using psychophysiological interaction (PPI) analyses: significantly activated ROIs during extinction learning and recall (also in interaction with cortisol) were entered as seed regions (VOI with a 5-mm sphere around the peak voxel).

## Results

### Day 1: Fear Acquisition

Higher SCRs were present for both CS+ compared with CS- (main effect CS:  $F_{(1,53)}=4.05, P=.049, \eta^2_p=.071$ ; supplementary Figure 1). In addition, successful fear acquisition was indicated by an increased activation of the fear network for both CS+ compared with CS- observed in the bilateral amygdala, insula, dACC, and right hippocampus (Table 1). Whole-brain analyses confirmed increased activation in response to both CS+ compared with CS- for the right insula (see supplementary Table 3 for whole-brain results in the critical contrasts). Importantly, no differences occurred between CS+G and CS+N in the predefined ROIs, the exploratory whole-brain analysis, or the SCRs.

### Day 2: Extinction

During extinction training, there was a significant decrease in SCRs ( $F_{(1,52)}=38.09, P<.001, \eta^2_p=.423$ ) from the first to the second half of extinction training for CS+G ( $F_{(1,52)}=40.42, P<.001, \eta^2_p=.437$ ), CS+N ( $F_{(1,52)}=22.42, P<.001, \eta^2_p=.301$ ), and CS- ( $F_{(1,52)}=13.47, P=.001, \eta^2_p=.206$ ; supplementary Figure 1). Additionally, there was a significant main effect of CS ( $F_{(1,47,76,40)}=17.90, P<.001,$

$\eta^2_p=.256$ ) across the entire phase of extinction training, indicating higher SCRs for CS+G ( $F_{(1,52)}=23.78, P<.001, \eta^2_p=.314$ ) and CS+N ( $F_{(1,52)}=16.81, P<.001, \eta^2_p=.244$ ) compared with CS- but not between CS+G and CS+N ( $F_{(1,52)}=2.11, P=.456, \eta^2_p=.039$ ). However, a significant CS  $\times$  half interaction ( $F_{(1,71,88,73)}=6.79, P=.003, \eta^2_p=.116$ ) indicated differential CS responding over time. Although SCRs for both CS+G ( $F_{(1,52)}=28.03, P<.001, \eta^2_p=.350$ ) and CS+N ( $F_{(1,52)}=19.03, P<.001, \eta^2_p=.268$ ) were significantly higher than for CS- during the first half of extinction training, this difference only persisted in the second half for CS+G ( $F_{(1,52)}=7.41, P=.026, \eta^2_p=.125$ ) but not for CS+N ( $F_{(1,52)}=5.84, P=.058, \eta^2_p=.101$ ) compared with CS- (supplementary Figure 1). However, there was a difference between CS+G and CS+N for neither the first ( $F_{(1,52)}=2.69, P=.320, \eta^2_p=.049$ ) nor the second half of extinction training ( $F_{(1,52)}=0.98, P>.999, \eta^2_p=.002$ ).

On the neural level, we observed a reduced activation of fear-related structures (bilateral amygdala, insula, dACC, and left hippocampus) for CS+G compared with CS- and reduced activation of the left insula, dACC, and vmPFC for CS+N compared with CS- in the second half compared with the first half of extinction training (see supplementary Tables 4 and 5 for the ROI and whole-brain results for the extinction contrasts).

Importantly, in the contrast CS+G minus CS+N, increased activation was found in the left amygdala during the first half and in the right insula during the second half of extinction training. Complementary, the vmPFC was more strongly deactivated for the CS+G compared with CS+N during the first half of extinction training (Table 1; Figure 2; see supplementary Table 3 for exploratory whole-brain results). No significant results were found in the performed PPIs ( $P>.05$ ).

### Day 3: Recall

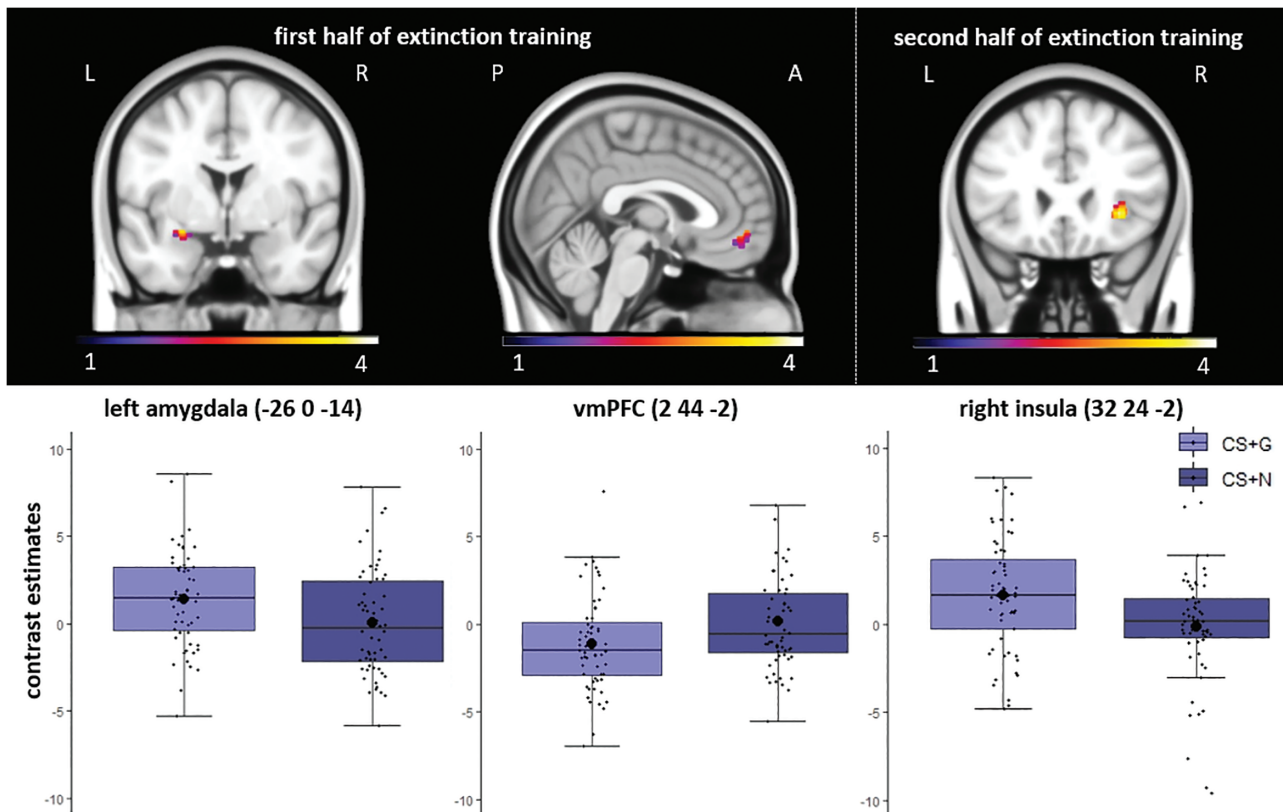
A significant main effect of CS ( $F_{(1,58,82,09)}=9.65, P=.001, \eta^2_p=.157$ ) indicated differences in SCRs during recall for CS+G ( $F_{(1,52)}=9.83,$

**Table 1.** Peak-voxel statistics and localizations for the contrast CS+ vs CS- for (a) fear acquisition training and the contrast CS+G vs CS+N (directions of the contrasts are marked) during (b) early and (c) late extinction training as well as during (d) recall. Similarly, (e) cortisol effects on recall are reported.

Contrast	Structure	Cluster size	x	y	z	T <sub>max</sub>	P <sub>corr</sub>
(a) Fear acquisition training							
CS+ > CS-	L amygdala	52	-18	-6	-12	3.33	.039
	R amygdala	68	18	-10	-16	3.33	.044
	L insula	616	-34	18	6	4.73	.003
	R insula	333	30	22	0	5.72	<.001
	dACC	81	4	16	38	4.45	.001
	R hippocampus	87	16	-10	-18	3.77	.029
(b) Early extinction training							
CS+G > CS+N	L amygdala	33	-26	0	-14	3.35	.041
CS+G < CS+N	vmPFC	70	2	44	-2	3.46	.009
(c) Late extinction training							
CS+G > CS+N	R insula	106	32	24	-2	3.93	.032
(d) Recall							
CS+G < CS+N	L amygdala	74	-22	-10	-14	3.39	.033
	R PHG	34	30	0	-34	3.67	.032
(e) Cortisol effects for CS+G > CS+N during recall							
Cortisol > placebo	L insula	81	-34	-6	10	3.84	.038
	L amygdala	68	-22	-10	-12	3.22	.050

Abbreviations: CS+, reinforced conditioned stimulus; CS-, non-reinforced conditioned stimulus; CS+G, generalized extinguished conditioned stimulus; CS+N, non-generalized extinguished conditioned stimulus; dACC, dorsal anterior cingulate cortex; L, left; PHG, parahippocampal gyrus; R, right; vmPFC, ventromedial prefrontal cortex.

The significance threshold was set to  $P \leq .05$  (family-wise error-corrected for small volume correction). All coordinates (x, y, z) are given in MNI space.



**Figure 2.** Differential neural responding for the contrast generalized extinguished conditioned stimulus (CS+G) minus non-generalized extinguished conditioned stimulus (CS+N) during extinction training. The slices were selected according to peak voxels of the activated ROI: (a) left amygdala, (b) vmPFC, and (c) right insula. Data are presented on the standard MNI brain template and thresholded to  $T \geq 1$  (see color bar for exact T values). Boxplots represent contrast estimates of the respective peak voxel (with black dots representing the mean). Abbreviations: A, anterior; L, left; P, posterior; R, right. During the first half of extinction training, increased activation of the left amygdala accompanied by decreased activation of the ventromedial prefrontal cortex (vmPFC) was observed for CS+G minus CS+N. In the second half of extinction training, right insula activation was increased for CS+G compared with CS+N.

$P = .008$ ,  $\eta^2_p = .159$ ) and CS+N ( $F_{(1,52)} = 13.72$ ,  $P = .002$ ,  $\eta^2_p = .209$ ) compared with CS- but not for the critical comparison of CS+G and CS+N ( $F_{(1,52)} = 0.12$ ,  $P > .999$ ,  $\eta^2_p = .002$ ). However, exploratory analyses revealed significantly higher SCRs for CS+N compared with CS- ( $F_{(1,52)} = 6.19$ ,  $P = .048$ ,  $\eta^2_p = .106$ ) for the modified size in the first trial but not for CS+G compared with CS- ( $F_{(1,52)} = 3.73$ ,  $P = .176$ ,  $\eta^2_p = .067$ ; [supplementary Figure 1](#)). Nevertheless, the direct comparison of CS+G and CS+N in SCRs revealed no significant difference ( $F_{(1,52)} = 0.23$ ,  $P > .999$ ,  $\eta^2_p = .004$ ).

Crucially, activations in the left amygdala and right parahippocampal gyrus were decreased for CS+G compared with CS+N for the original and modified stimuli, partially supporting our first hypothesis on the neural basis of extinction generalization ([Table 1](#); [Figure 3](#)). Psychophysiological interactions indicated that the deactivation of the parahippocampal gyrus was connected to increased activation in the right hippocampus. All other PPIs were nonsignificant ( $P > .05$ ). Exploratory whole-brain analyses and the comparison of CS+G and CS+N for the modified size only revealed no significant results ( $P > .05$ ).

### Day 3: Salivary Cortisol

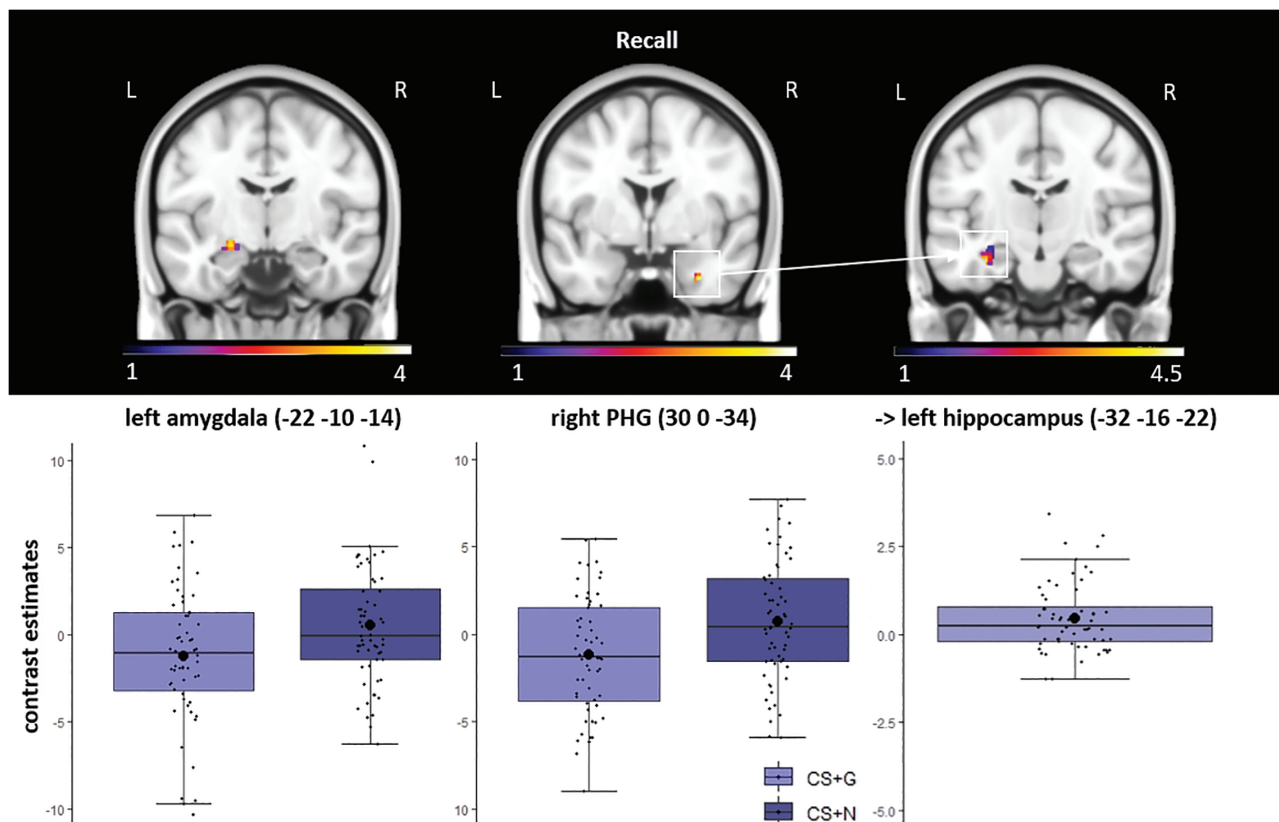
In addition to significant main effects of time ( $F_{(1,27,69,66)} = 79.32$ ,  $P < .001$ ,  $\eta^2_p = .591$ ) and group ( $F_{(1,55)} = 35.85$ ,  $P < .001$ ,  $\eta^2_p = .395$ ), enhanced cortisol levels in the cortisol group were indicated by a significant time  $\times$  group interaction ( $F_{(1,27,69,66)} = 105.00$ ,  $P < .001$ ,  $\eta^2_p = .656$ ). Planned comparisons revealed higher cortisol levels in

the cortisol group compared with the placebo group 30 minutes ( $F_{(1,55)} = 61.33$ ,  $P < .001$ ,  $\eta^2_p = .527$ ) and 60 minutes after tablet intake ( $F_{(1,55)} = 51.93$ ,  $P < .001$ ,  $\eta^2_p = .486$ ), indicating a successful cortisol manipulation ([supplementary Table 1](#)).

### Day 3: Cortisol Effects on Recall

There was no significant main or interaction effect of cortisol for SCRs ( $P_s > .156$ ; [supplementary Figure 2](#)).

According to our second hypothesis, we expect the difference between CS+G and CS+N to be higher in the cortisol compared with the placebo group as reflected in activation of the fear network. Indeed, the above-mentioned decreased activation of the left amygdala during recall for CS+G compared with CS+N interacted with cortisol: a decreased amygdala activation was observed for CS+G compared with CS+N in the placebo group but not in the cortisol group ([Figure 4](#)). A comparable pattern emerged for the left insula: while a decreased insula activation emerged for CS+G compared with CS+N in the placebo group, insula activation was increased for CS+G compared with CS+N in the cortisol group ([Figure 4](#)). Additionally, the left insula had stronger functional connections to the vmPFC in the placebo compared with the cortisol group in the contrast CS+G minus CS+N ([Figure 4](#)). Comparisons of the CS+G and CS+N with the CS- seem to confirm that the observed cortisol effects mainly rely on the CS+G: activation of the amygdala is enhanced in the cortisol group compared with the placebo group (see [supplementary Table 6](#)). Taken together, the cortisol results revealed no



**Figure 3.** Differential neural responding for generalized extinguished conditioned stimulus (CS+G) minus non-generalized extinguished conditioned stimulus (CS+N) during recall. The slices were selected according to peak voxels of the activated ROIs: (a) left amygdala and (b) right parahippocampal gyrus (PHG). Functional connectivity between the right PHG and (c) the left hippocampus ( $T=4.02$ ,  $P=.015$ ) is displayed. Data are presented on the standard MNI brain template and thresholded to  $T \geq 1$  (see color bar for exact  $T$  values). Boxplots represent contrast estimates of the respective peak voxel (with black dots representing the mean). Abbreviations: A, anterior; L, left; P, posterior; R, right. During recall, both the left amygdala and right PHG show a decreased activation to CS+G compared with CS+N.

support for our second hypothesis stating that extinction generalization might be able to overcome the detrimental effects of cortisol administration prior to recall based on the above reported neural findings.

Other PPIs were nonsignificant ( $P > .05$ ). Exploratory whole-brain analyses revealed no significance for the contrast CS+G compared with CS+N. For the contrast CS+G minus CS+N regarding the modified stimulus size, there were no significant results. Thus, pure generalization effects were not subject to a modulation by cortisol. Furthermore, no main effects or interactions of the factors cortisol and sex were present during fear acquisition training, extinction training, or recall ( $P > .05$ ). Whole-brain analyses also did not reveal any significant results for cortisol interactions ( $P > .05$ ).

## Discussion

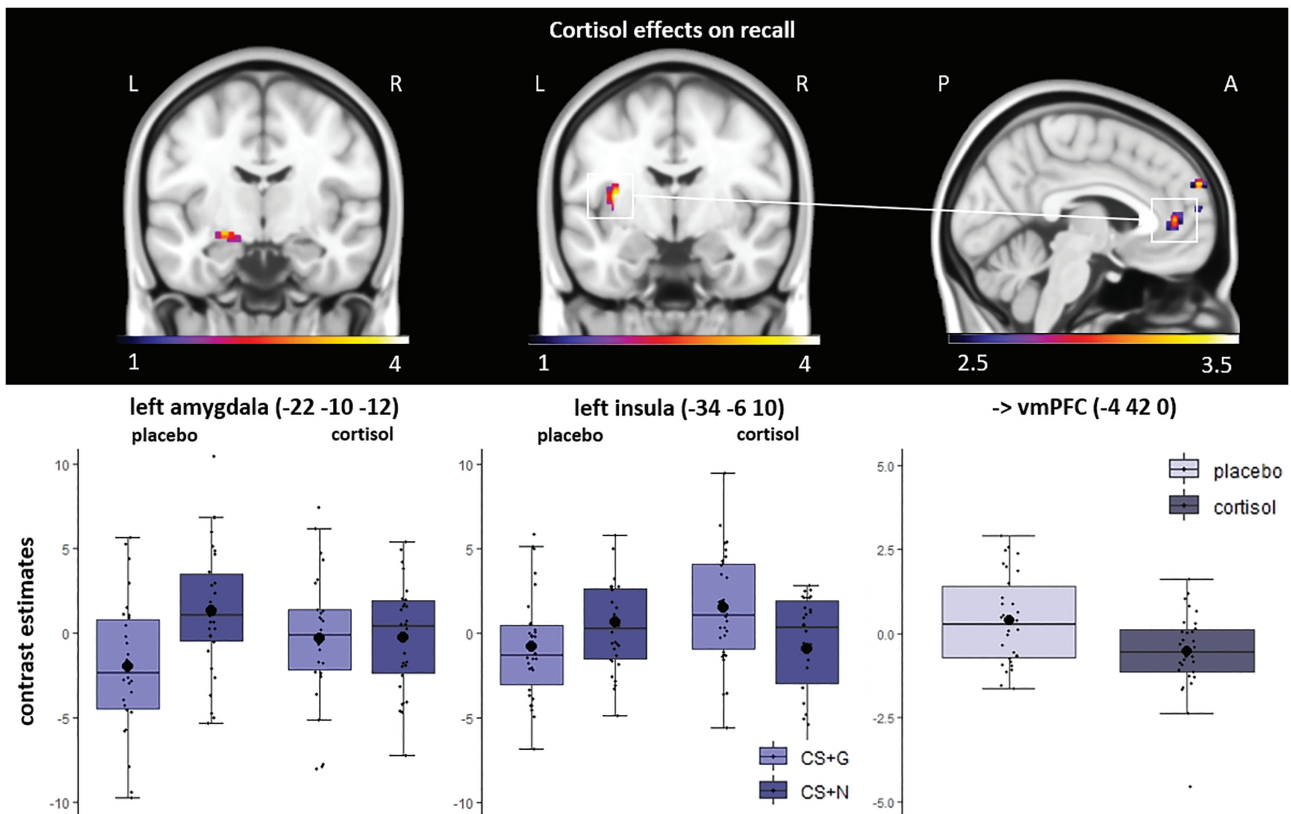
This study aimed to identify the neural correlates of extinction generalization and to characterize the effects of cortisol administration on extinction generalization recall for the first time.

Increased amygdala and insula activation as well as decreased vmPFC activation were observed for CS+G compared with CS+N during extinction training, possibly indicating a prolonged extinction learning (Figure 2). However, the comparison of CS+G and CS+N revealed nonsignificant results for the SCR analysis. Thus, alternative explanations for these altered neural processes should be considered. For example, the modified sizes of the CS+G might have elicited novelty and attentional effects

causing salience (Craske et al., 2018). Thus, processes other than fear-related signaling per se could underlie CS+G processing. However, a previous study investigating extinction generalization reported enhanced arousal during extinction training for the CS+G as reflected by SCRs (Waters et al., 2018). Yet we could not replicate this effect: slightly increased SCRs during the second half of extinction training for CS+G compared with CS- emerged, but the more critical comparison of CS+G and CS+N was not significant. However, both processes (i.e., novelty and fear-related signaling) might enhance attention towards the CS+G: due to increased novelty and attention, encoding of the CS+G during extinction training might be facilitated, consequently leading to an enhanced extinction memory recall 1 day later.

The altered neural processing of the CS+G compared with CS+N could also be a consequence of the fewer presentations of each size of the CS+G (twice) compared with the CS+N (8 times) during extinction training. However, a previous report indicated that incorporating multiple generalization stimuli during extinction training prolonged the extinction process, even if the same number of the original CS+G and CS+N was presented (Waters et al., 2018). Thus, the reported changes in neural processing more likely rely on the implementation of generalization stimuli presented in different sizes, probably enhancing attention towards the CS+G. In addition, balancing the total number of CS+G and CS+N yields the advantage of controlling learning effects across all CS: due to the balanced number of CS+G and CS+N presentations, differential learning effects cannot rely on differences in the number of presentations.





**Figure 4.** Cortisol effects on recall for the contrast generalized extinguished conditioned stimulus (CS+G) minus non-generalized extinguished conditioned stimulus (CS+N). The slices were selected according to peak voxels of the activated ROIs: (a) left amygdala and (b) left insula. Additionally, functional connectivity between the left insula and (c) the ventromedial prefrontal cortex (vmPFC) ( $T=2.72$ ,  $P=.046$ ) is displayed. Data is presented on the standard MNI brain template and thresholded to  $T \geq 1$  for activations and  $T \geq 2.5$  for functional connectivity (see color bar for exact  $T$  values). Boxplots represent contrast estimates of the respective peak voxel (with black dots representing the mean). Abbreviations: A, anterior; L, left; P, posterior; R, right. Increased differential neural responding was present for CS+G minus CS+N for the cortisol compared with the placebo group.

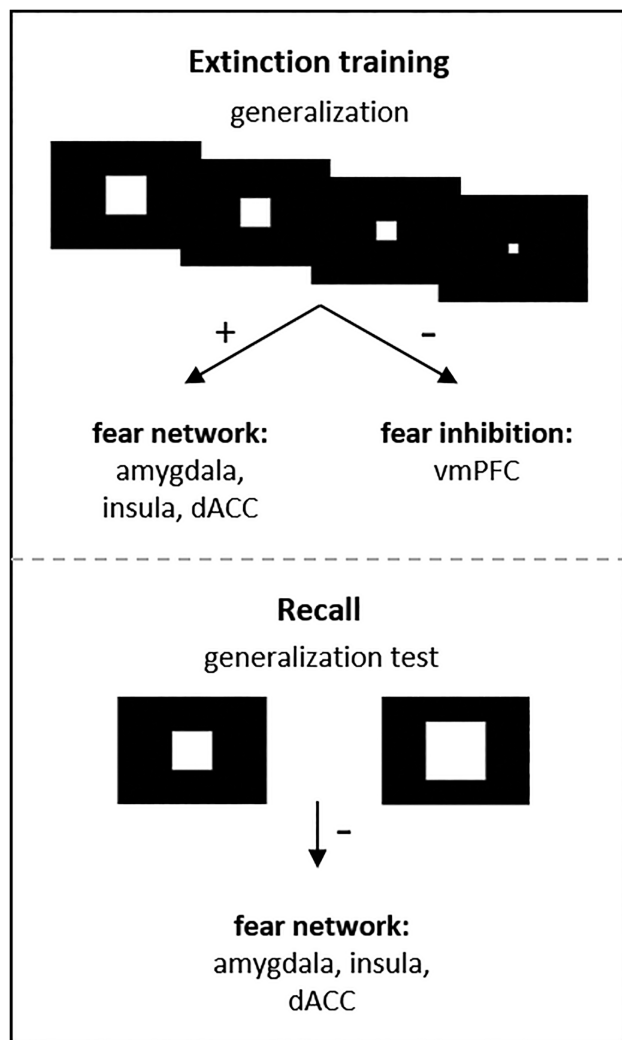
According to attentional-associative models (Pearce and Hall, 1980; Dunsmoor and Schmajuk, 2009), variations in CS might enhance the prediction error for the occurrence of the negative outcome as it reallocates attentional resources towards the new variations of the generalized extinction stimulus (Roesch et al., 2012). Higher prediction errors should consequently result in enhanced extinction learning for both, the original stimulus, and even formerly unrepresented generalization stimuli (Figure 5).

During recall, although not reflected in SCRs, diminished amygdala activation was present for the CS+G compared with CS+N, possibly depicting a neural correlate of extinction generalization in line with our first hypothesis. Though the amygdala is commonly regarded as vital for fear learning, recent meta-analyses of fMRI findings in the area of fear conditioning (Fullana et al., 2016, 2018) challenged this notion. Possible explanations encompass the missing activation of the survival-defensive circuits in human fear conditioning paradigms due to ethical limitations or the comparison of threat-related stimuli (CS+) to safety stimuli (CS-; Fullana et al., 2019). The latter explanation, however, cannot account for differential activation patterns within the critical contrast CS+G compared with CS+N. Although a reduced amygdala activation for generalized extinguished stimuli was already observed after 1 day, other studies point toward stronger effects after multiple weeks (Rowe and Craske, 1998; Shiban et al., 2015).

Alternative, but not necessarily mutually exclusive, explanations might also account for the observed effects on recall. For example, increased salience or novelty during extinction training could enhance attention and alter extinction learning for the CS+G. Thus, increased attention toward the CS+G could result in enhanced extinction learning despite of continued fear-related processing. In addition, a more general (declarative) emotional memory recall (Dunsmoor and Kroes, 2019) through an interplay of amygdala, hippocampal (Phelps, 2004; Richter-Levin, 2004; Phelps and LeDoux, 2005), and prefrontal regions (Preston and Eichenbaum, 2013) instead of a more specific fear (or extinction) recall could be assumed. In the absence of physiological arousal reflected in SCRs, these activations might indicate the formation and recall of an emotional memory trace in the amygdala and parahippocampal gyrus (PHG) and its functional connectivity to the hippocampus.

Decreased PHG activation to the CS+G might reflect less recollection of source memory compared with the CS+N (Stevenson et al., 2020). In accordance, the functional connectivity of the increased PHG activation to the decreased hippocampal involvement might argue for enhanced pattern completion in the hippocampus (Treves and Rolls, 1994; Figure 5). However, no enhanced activation of the hippocampus per se was present during extinction learning and recall. Consequently, the distinction between the hippocampal subregions with higher magnetic field strength might constitute a promising approach to further characterize extinction generalization mechanisms.





**Figure 5.** Neural model of extinction generalization based on the current findings. Extinction generalization results in prolonged fear expression reflected in amygdala and insula activation and delayed safety signaling in the ventromedial prefrontal cortex (vmPFC) during extinction training. Successful extinction generalization is mirrored in reduced amygdala activation and thus return of fear. Plus symbols indicate upregulation while minus symbols indicate downregulation of the respective area.

Increased variability and limited predictability due to the inclusion of multiple similar stimuli during extinction generalization training might violate US expectancy and therefore increase prediction errors (Figure 5). Since prediction errors are assumed to act as one of the main drivers of exposure therapy success (Vervliet et al., 2013b; Craske et al., 2018), increasing the prediction error should strengthen the extinction memory trace and facilitate generalization processes. Previously, neural correlates of prediction errors and attention-modulated representations have been identified simulating an attentional-associative model in a fear conditioning study manipulating the reinforcement rate (Dunsmoor and Schmajuk, 2009). In this study, amygdala and ACC activation most likely reflected prediction error signaling, whereas dorsolateral PFC activation most likely mirrored attention-modulated representations.

In line with attentional-associative models, extinction generalization appears to rely on prolonged activation of the amygdala and insula accompanied by higher physiological arousal

(Waters et al., 2018) during extinction learning in which prediction errors can be expected due to the varying presentation of 1 formerly reinforced CS. Enhanced prediction error signaling might be involved in a multitude of findings in extinction generalization research, including extinction generalization using a peak-stimulus (Struyf et al., 2018) and the absence of extinction generalization including 1 (Wong and Lovibond, 2020) or multiple (Zbozinek and Craske, 2018) generalization stimuli, but not the original CS. However, it should be taken into consideration that reduced fear recall and enhanced extinction recall can both account for the observed decreased activation of the amygdala during recall since the amygdala comprises both fear and safety neurons (Genud-Gabai et al., 2013).

Partially in line with previous studies (Kinner et al., 2016, 2018; for a review, see Meir Drexler et al., 2019), cortisol administration prior to recall increased amygdala and insula activation that was negatively connected to the vmPFC in the cortisol group (Figure 4). Thus, the advantageous effects of extinction generalization on neural activation patterns appeared to be blocked by cortisol administration, possibly representing a model for stress-induced return of fear observed in the context of anxiety disorders (Jacobs and Nadel, 1985; Francis et al., 2012). Cortisol effects on the BOLD level were not accompanied by similar effects on the SCR level. As SCRs might capture a more contingency awareness-related aspect (Tabbert et al., 2011), neural activations might be more sensitive than SCRs to detect extinction generalization effects during extinction training and recall. Taken together and in contrast to our second hypothesis, extinction generalization did not seem to counteract the detrimental effects of cortisol administration prior to recall.

One explanation for the neural findings might be a reduced hippocampal activation in the aftermath of stress (Meir Drexler et al., 2019). Generalization effects are assumed to mainly rely on pattern separation and pattern completion in the hippocampus (Lissek et al., 2014). Consequently, the ability to retrieve the generalized extinction memory trace appears to be limited in the face of stress, comparable with the negative effects of stress hormones on declarative memory retrieval (de Quervain et al., 2000, 2017; Smeets et al., 2008; Smeets, 2011; Wolf, 2017). In addition, emotional learning processes that rely on interactions between the amygdala and hippocampus might also partially account for these findings; cortisol administration prior to recall appears to deteriorate memory retrieval (Wolf, 2009) similar to a blocked extinction recall after cortisol administration (Kinner et al., 2016, 2018). Even though extinction memory recall in the placebo group seemed to be enhanced, cortisol impaired extinction recall irrespective of former extinction generalization.

Although cortisol administration serves as a mechanistic model for stress, stress is not only characterized by a release of cortisol but also by additional bio-psycho-social factors (McEwen, 2007; Joëls and Baram, 2009). Thus, future studies should also consider stress paradigms and their possible impact on the recall of generalized extinction memories. Nevertheless, cortisol administration allows for mechanistic investigations of cortisol effects on learning and memory processes as realized before (for reviews, see Het et al., 2005; Wolf, 2009; Meir Drexler et al., 2019). These cortisol effects on learning and memory processes are key to our understanding of the pathogenesis and treatment of various mental disorders, including anxiety disorders and posttraumatic stress disorder (Bentz et al., 2010; de Quervain et al., 2017). For example, cortisol itself might comprise the potential to overcome stress-induced relapses: cortisol administered prior to extinction training is supposed to strengthen the extinction memory trace, which in turn counteracts the return of fear

in healthy participants as well as patients with anxiety disorders (Soravia et al., 2006, 2014; Bentz et al., 2010, 2013; Brueckner et al., 2019). Additionally, further strategies enhancing the recall of the extinction memory trace, for example deepened extinction or occasional reinforced extinction (Craske et al., 2014, 2018), could be applied to reduce the risk of stress-induced relapses, potentially in combination with extinction generalization.

The application of more naturalistic stimuli and inclusion of clinical populations would considerably extend the findings of the present study to real-life therapeutic settings. An investigation on spider-phobic patients already indicated successful stimulus-based extinction generalization being superior to context-based extinction generalization and their combination (Shiban et al., 2015). Importantly, the transfer of stimulus-based extinction generalization to various contexts can be considered particularly advantageous, since it counteracts renewal (Vervliet et al., 2013a; Andreatta et al., 2015; Podlesnik and Miranda-Dukoski, 2015), 1 major source of return of fear and consequently relapses.

Although sex differences in cortisol effects on extinction recall have been reported previously (Merz et al., 2018b; Velasco et al., 2019), this study on cortisol effects on extinction generalization recall failed to reveal any differences. Despite missing evidence for sex differences in extinction generalization processes, the lack of control for menstrual cycle phase may also account for this finding. Hence, further studies focusing on sex differences and the impact of sex hormones such as estrogens on extinction generalization processes are required.

In conclusion, stimulus-based extinction generalization increased amygdala and insula activation during extinction learning while decreasing vmPFC activation relative to classical extinction learning. Thus, extinction generalization might enhance arousal and increase novelty or salience during extinction learning. During recall, a decreased activation of the amygdala and PHG accompanied by increased functional connectivity to the hippocampus was observed, which might point towards reduced fear expression or less emotional memory recall in response to the CS+G compared with CS+N. Analogous to the effects observed for extinction learning, this effect could also be partially explained by novelty or salience effects; due to greater variability during extinction learning, novelty decreases for the CS+G during recall. Despite missing effects on SCRs, these results in combination might constitute a first hint at positive effects of extinction generalization on fear-related structures. However, this neural pattern was blocked by cortisol administration; amygdala and insula activation was decreased and a reduced functional connectivity to the vmPFC for the CS+G compared with CS+N emerged in the placebo group, whereas no difference between CS+G and CS+N was observed in the cortisol group. Thus, the implementation of generalization stimuli might foster extinction learning, whereas cortisol appears to counteract these effects during recall, corroborating previous reports of detrimental cortisol effects on the recall of (declarative) emotional memories in general.

## Acknowledgments

We gratefully acknowledge the help of Annegret Last, Alexander Böckmann, Beate Krauß, Anna Metzler, and Dennis Pomrehn during data collection and recruitment of participants. Moreover, we thank Tobias Otto and PHILIPS Germany for technical support and Jaël Katharina Caviola, Leja Zoe Korneffel, Ronja Canjara Leisch, and Pia Katharina Strater for language editing.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) within the SFB 1280 Extinction Learning (grant no. 316803389 - SFB1280; project A09 [OTW, CJM]). The DFG had no role in study design, collection, analysis and interpretation of data, writing of the manuscript, or in the decision to submit the paper for publication.

## Statement of Interest

None.

## References

- Andreatta M, Leombruni E, Glotzbach-Schoon E, Pauli P, Mühlberger A (2015) Generalization of contextual fear in humans. *Behav Ther* 46:583–596.
- Asok A, Kandel ER, Rayman JB (2019) The neurobiology of fear generalization. *Front Behav Neurosci* 12:1–15.
- Benedek M, Kaernbach C (2010a) A continuous measure of phasic electrodermal activity. *J Neurosci Methods* 190:80–91.
- Benedek M, Kaernbach C (2010b) Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47:647–658.
- Bennett M, Vervoort E, Boddez Y, Hermans D, Baeyens F (2015) Perceptual and conceptual similarities facilitate the generalization of instructed fear. *J Behav Ther Exp Psychiatry* 48:149–155.
- Bentz D, Michael T, de Quervain DJ, Wilhelm FH (2010) Enhancing exposure therapy for anxiety disorders with glucocorticoids: from basic mechanisms of emotional learning to clinical applications. *J Anxiety Disord* 24:223–230.
- Bentz D, Michael T, Wilhelm FH, Hartmann FR, Kunz S, von Rohr IR, de Quervain DJ (2013) Influence of stress on fear memory processes in an aversive differential conditioning paradigm in humans. *Psychoneuroendocrinology* 38:1186–1197.
- Bouton ME (2002) Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry* 52:976–986.
- Bouton ME (2004) Context and behavioral processes in extinction. *Learn Mem* 11:485–494.
- Bouton ME, Todd TP (2014) A fundamental role for context in instrumental learning and extinction. *Behav Processes* 104:13–19.
- Brueckner AH, Lass-Hennemann J, Wilhelm FH, Ferreira de Sá DS, Michael T (2019) Cortisol administration after extinction in a fear-conditioning paradigm with traumatic film clips prevents return of fear. *Transl Psychiatry* 9:128.
- Chung S, Son GH, Kim K (2011) Circadian rhythm of adrenal glucocorticoid: its regulation and clinical implications. *Biochim Biophys Acta - Mol Basis Dis* 1812:581–591.
- Craske MG, Hermans D, Vansteenwegen D eds. (2006) *Fear and learning: from basic processes to clinical implications*. Washington, DC: American Psychological Association.
- Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B (2014) Maximizing exposure therapy: an inhibitory learning approach. *Behav Res Ther* 58:10–23.
- Craske MG, Hermans D, Vervliet B (2018) State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philos Trans R Soc B Biol Sci* 373:20170025.
- de Quervain DJ, Roozendaal B, Nitsch RM, McGaugh JL, Hock C (2000) Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nat Neurosci* 3:313–314.

- de Quervain D, Schwabe L, Roozendaal B (2017) Stress, glucocorticoids and memory: implications for treating fear-related disorders. *Nat Rev Neurosci* 18:7–19.
- de Quervain D, Wolf OT, Roozendaal B (2019) Glucocorticoid-induced enhancement of extinction from animal models to clinical trials. *Psychopharmacology (Berl)* 236:183–199.
- de Voogd LD, Murray YPJ, Barte RM, van der Heide A, Fernández G, Doeller CF, Hermans EJ (2020) The role of hippocampal spatial representations in contextualization and generalization of fear: fear memory contextualization. *Neuroimage* 206:116308.
- Deschaux O, Zheng X, Lavigne J, Nachon O, Cleren C, Moreau JL, Garcia R (2013) Post-extinction fluoxetine treatment prevents stress-induced reemergence of extinguished fear. *Psychopharmacology (Berl)* 225:209–216.
- Dunsmoor JE, Kroes MCW (2019) Episodic memory and Pavlovian conditioning: ships passing in the night. *Curr Opin Behav Sci* 26:32–39.
- Dunsmoor JE, Martin A, LaBar KS (2012) Role of conceptual knowledge in learning and retention of conditioned fear. *Biol Psychol* 89:300–305.
- Dunsmoor JE, Murphy GL (2015) Categories, concepts, and conditioning: how humans generalize fear. *Trends Cogn Sci* 19:73–77.
- Dunsmoor J, Schmajuk N (2009) Interpreting patterns of brain activation in human fear conditioning with an attentional-associative learning model. *Behav Neurosci* 123:851–855.
- Dunsmoor JE, White AJ, LaBar KS (2011) Conceptual similarity promotes generalization of higher order fear learning. *Learn Mem* 18:156–160.
- Faul F, Erdfelder E, Lang AG, Buchner A (2007) G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39:175–191.
- Forcadell E, Torrents-Rodas D, Vervliet B, Leiva D, Tortella-Feliu M, Fullana MA (2017) Does fear extinction in the laboratory predict outcomes of exposure therapy? A treatment analog study. *Int J Psychophysiol* 121:63–71.
- Francis JL, Moitra E, Dyck I, Keller MB (2012) The impact of stressful life events on relapse of generalized anxiety disorder. *Depress Anxiety* 29:386–391.
- Fullana MA, Harrison BJ, Soriano-Mas C, Vervliet B, Cardoner N, Àvila-Parcet A, Radua J (2016) Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21:500–508.
- Fullana MA, Albajes-Eizaguirre A, Soriano-Mas C, Vervliet B, Cardoner N, Benet O, Radua J, Harrison BJ (2018) Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants. *Neurosci Biobehav Rev* 88:16–25.
- Fullana MA, Albajes-Eizaguirre A, Soriano-Mas C, Vervliet B, Cardoner N, Benet O, Radua J, Harrison BJ (2019) Amygdala where art thou? *Neurosci Biobehav Rev* 102:430–431.
- Genud-Gabai R, Klavir O, Paz R (2013) Safety signals in the primate amygdala. *J Neurosci* 33:17986–17994.
- Hamacher-Dang TC, Uengoer M, Wolf OT (2013) Stress impairs retrieval of extinguished and unextinguished associations in a predictive learning task. *Neurobiol Learn Mem* 104:1–8.
- Hammoud MZ, Foa EB, Milad MR (2020) Oestradiol, threat conditioning and extinction, post-traumatic stress disorder, and prolonged exposure therapy: a common link. *J Neuroendocrinol* 32:e12800.
- Het S, Ramlow G, Wolf OT (2005) A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology* 30:771–784.
- Holt DJ, Boeke EA, Wolthusen RP, Nasr S, Milad MR, Tootell RB (2014) A parametric study of fear generalization to faces and non-face objects: relationship to discrimination thresholds. *Front Hum Neurosci* 8:624.
- Hwang MJ, Zsido RG, Song H, Pace-Schott EF, Miller KK, Lebron-Milad K, Marin MF, Milad MR (2015) Contribution of estradiol levels and hormonal contraceptives to sex differences within the fear network during fear conditioning and extinction. *BMC Psychiatry* 15:295.
- Jacobs WJ, Nadel L (1985) Stress-induced recovery of fears and phobias. *Psychol Rev* 92:512–531.
- Joëls M, Baram TZ (2009) The neuro-symphony of stress. *Nat Rev Neurosci* 10:459–466.
- Kinner VL, Merz CJ, Lissek S, Wolf OT (2016) Cortisol disrupts the neural correlates of extinction recall. *Neuroimage* 133:233–243.
- Kinner VL, Wolf OT, Merz CJ (2018) Cortisol increases the return of fear by strengthening amygdala signaling in men. *Psychoneuroendocrinology* 91:79–85.
- Lange I, Goossens L, Michielse S, Bakker J, Vervliet B, Marcelis M, Wichers M, van Os J, van Amelsvoort T, Schruers K (2020) Neural responses during extinction learning predict exposure therapy outcome in phobia: results from a randomized-controlled trial. *Neuropsychopharmacology* 45:534–541.
- Lissek S, Powers AS, McClure EB, Phelps EA, Woldehawariat G, Grillon C, Pine DS (2005) Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav Res Ther* 43:1391–1424.
- Lissek S, Bradford DE, Alvarez RP, Burton P, Espensen-Sturges T, Reynolds RC, Grillon C (2014) Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. *Soc Cogn Affect Neurosci* 9:1134–1142.
- McEwen BS (2007) Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol Rev* 87:873–904.
- Mechias ML, Etkin A, Kalisch R (2010) A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. *Neuroimage* 49:1760–1768.
- Meir Drexler S, Merz CJ, Jentsch VL, Wolf OT (2019) How stress and glucocorticoids timing-dependently affect extinction and relapse. *Neurosci Biobehav Rev* 98:145–153.
- Merz CJ, Hamacher-Dang TC, Stark R, Wolf OT, Hermann A (2018a) Neural underpinnings of cortisol effects on fear extinction. *Neuropsychopharmacology* 43:384–392.
- Merz CJ, Kinner VL, Wolf OT (2018b) Let's talk about sex ... differences in human fear conditioning. *Curr Opin Behav Sci* 23:7–12.
- Merz CJ, Wolf OT (2017) Sex differences in stress effects on emotional learning. *J Neurosci Res* 95:93–105.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Pappens M, Schroyen M, Van den Bergh O, Van Diest I (2015) Retention of perceptual generalization of fear extinction. *Int J Psychophysiol* 98:520–528.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552.
- Penny W, Friston K, Ashburner J, Kiebel S, Nichols T (2007) Statistical parametric mapping (Penny W, Friston K, Ashburner J, Kiebel S, Nichols T, eds). San Diego, CA: Elsevier.
- Phelps EA (2004) Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr Opin Neurobiol* 14:198–202.



- Phelps EA, LeDoux JE (2005) Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48:175–187.
- Podlesnik CA, Miranda-Dukoski L (2015) Stimulus generalization and operant context renewal. *Behav Processes* 119:93–98.
- Preston AR, Eichenbaum H (2013) Interplay of hippocampus and prefrontal cortex in memory. *Curr Biol* 23:R764–R773.
- Raio CM, Brignoni-Perez E, Goldman R, Phelps EA (2014) Acute stress impairs the retrieval of extinction memory in humans. *Neurobiol Learn Mem* 112:212–221.
- Richter-Levin G (2004) The amygdala, the hippocampus, and emotional modulation of memory. *Neuroscientist* 10:31–39.
- Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G (2012) Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *Eur J Neurosci* 35:1190–1200.
- Rowe MK, Craske MG (1998) Effects of varied-stimulus exposure training on fear reduction and return of fear. *Behav Res Ther* 36:719–734.
- Scheveneels S, Boddez Y, Vervliet B, Hermans D (2016) The validity of laboratory-based treatment research: bridging the gap between fear extinction and exposure treatment. *Behav Res Ther* 86:87–94.
- Schiller D, Delgado MR (2010) Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn Sci* 14:268–276.
- Shiban Y, Schelhorn I, Pauli P, Mühlberger A (2015) Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. *Behav Res Ther* 71:45–53.
- Shields GS, Sazma MA, McCullough AM, Yonelinas AP (2017) The effects of acute stress on episodic memory: a meta-analysis and integrative review. *Psychol Bull* 143:636–675.
- Smeets T (2011) Acute stress impairs memory retrieval independent of time of day. *Psychoneuroendocrinology* 36:495–501.
- Smeets T, Otgaar H, Candel I, Wolf OT (2008) True or false? Memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology* 33:1378–1386.
- Soravia LM, Heinrichs M, Aerni A, Maroni C, Schelling G, Ehlert U, Roozendaal B, de Quervain DJ (2006) Glucocorticoids reduce phobic fear in humans. *Proc Natl Acad Sci U S A* 103:5585–5590.
- Soravia LM, Heinrichs M, Winzeler L, Fisler M, Schmitt W, Horn H, Dierks T, Strik W, Hofmann SG, de Quervain DJ (2014) Glucocorticoids enhance in vivo exposure-based therapy of spider phobia. *Depress Anxiety* 31:429–435.
- Starita F, Kroes MCW, Davachi L, Phelps EA, Dunsmoor JE (2019) Threat learning promotes generalization of episodic memory. *J Exp Psychol Gen* 148:1426–1434.
- Stevenson RF, Reagh ZM, Chun AP, Murray EA, Yassa MA (2020) Pattern separation and source memory engage distinct hippocampal and neocortical regions during retrieval. *J Neurosci* 40:843–851.
- Struyf D, Hermans D, Vervliet B (2018) Maximizing the generalization of fear extinction: exposures to a peak generalization stimulus. *Behav Res Ther* 111:1–8.
- Struyf D, Zaman J, Hermans D, Vervliet B (2017) Gradients of fear: how perception influences fear generalization. *Behav Res Ther* 93:116–122.
- Struyf D, Zaman J, Vervliet B, Van Diest I (2015) Perceptual discrimination in fear generalization: mechanistic and clinical implications. *Neurosci Biobehav Rev* 59:201–207.
- Tabbert K, Merz CJ, Klucken T, Schweckendiek J, Vaitl D, Wolf OT, Stark R (2011) Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. *Soc Cogn Affect Neurosci* 6:495–506.
- Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. *Hippocampus* 4:374–391.
- Velasco ER, Florido A, Milad MR, Andero R (2019) Sex differences in fear extinction. *Neurosci Biobehav Rev* 103:81–108.
- Vervliet B, Vansteenwegen D, Eelen P (2006) Generalization gradients for acquisition and extinction in human contingency learning. *Exp Psychol* 53:132–142.
- Vervliet B, Baeyens F, Van den Bergh O, Hermans D (2013a) Extinction, generalization, and return of fear: a critical review of renewal research in humans. *Biol Psychol* 92:51–58.
- Vervliet B, Craske MG, Hermans D (2013b) Fear extinction and relapse: state of the art. *Annu Rev Clin Psychol* 9:215–248.
- Vervoort E, Vervliet B, Bennett M, Baeyens F (2014) Generalization of human fear acquisition and extinction within a novel arbitrary stimulus category. *Plos One* 9:e96569.
- Waters AM, Kershaw R, Lipp OV (2018) Multiple fear-related stimuli enhance physiological arousal during extinction and reduce physiological arousal to novel stimuli and the threat conditioned stimulus. *Behav Res Ther* 106:28–36.
- Wolf OT (2009) Stress and memory in humans: twelve years of progress? *Brain Res* 1293:142–154.
- Wolf OT (2017) Stress and memory retrieval: mechanisms and consequences. *Curr Opin Behav Sci* 14:40–46.
- Wong AHK, Lovibond PF (2020) Generalization of extinction of a generalization stimulus in fear learning. *Behav Res Ther* 125:103535.
- Xu L, Su H, Xie X, Yan P, Li J, Zheng X (2018) The topological properties of stimuli influence fear generalization and extinction in humans. *Front Psychol* 9:1–10.
- Zaman J, Struyf D, Ceulemans E, Beckers T, Vervliet B (2019) Probing the role of perception in fear generalization. *Sci Rep* 9:10026.
- Zbozinek TD, Craske MG (2018) Pavlovian extinction of fear with the original conditional stimulus, a generalization stimulus, or multiple generalization stimuli. *Behav Res Ther* 107:64–75.