OXFORD

Social Cognitive and Affective Neuroscience, 2020, 625–634

doi: 10.1093/scan/nsaa074
Advance Access Publication Date: 08 June 2020
Original Manuscript

Dissociable neural signatures of passive extinction and instrumental control over threatening events

Nadine Wanke and Lars Schwabe

Department of Cognitive Psychology, Universität Hamburg, Hamburg, Germany

Correspondence should be addressed to Lars Schwabe, PhD, Universität Hamburg, Department of Cognitive Psychology, 20146 Hamburg, Germany. E-mail: Lars.Schwabe@uni-hamburg.de

Abstract

Aberrant fear learning processes are assumed to be a key factor in the pathogenesis of anxiety disorders. Thus, effective behavioral interventions to reduce dysfunctional fear responding are needed. Beyond passive extinction learning, instrumental control over threatening events is thought to diminish fear. However, the neural mechanisms underlying instrumental control—and to what extent these differ from extinction—are not well understood. We therefore contrasted the neural signatures of instrumental control and passive extinction using an aversive learning task, relative to a control condition. Participants ($n = 64$) could either learn to exert instrumental control over electric shocks, received a yoked number and sequence of shocks without instrumental control or did not receive any shocks. While both passive extinction and instrumental control reduced threat-related skin conductance responses (SCRs) relative to pre-extinction/control, instrumental control resulted in a significantly more pronounced decrease of SCRs. Instrumental control was further linked to decreased striatal activation and increased cross talk of the ventromedial prefrontal cortex (vmPFC) with the amygdala, whereas passive extinction was associated with increased vmPFC activation. Our findings demonstrate that instrumental learning processes may shape Pavlovian fear responses and that the neural underpinnings of instrumental control are critically distinct from those of passive extinction learning.

Key words: instrumental learning; yoked extinction; fear conditioning; striatum; vmPFC

Introduction

During Pavlovian fear conditioning, a neutral stimulus is repeatedly paired with a fear-eliciting unconditioned stimulus (UCS) until the initially neutral—now conditioned—stimulus (CS) alone is capable of triggering a fear response similar to the one produced by the UCS (Pavlov, 2010). This basic form of learning, preserved across species (Maren, 2001; Calhoon and Tye, 2015), is highly adaptive as it helps the organism to avoid current or future harm. Aberrant fear learning, however, may be dysfunctional and has been implicated in the pathogenesis of anxiety disorders or post-traumatic stress disorder (PTSD; Rosen and Schulkin, 1998; Rauch *et al.*, 2006; Milad *et al.*, 2009).

The most common therapeutic approach to treat these fear-related disorders, exposure therapy, is based on the principle of fear extinction and involves repeated exposure to the threatening stimulus in the absence of the aversive outcome, thus promoting new learning that the threatening stimulus is now safe (Maren and Quirk, 2004). On a neural level, extinction learning involves mainly the ventromedial prefrontal cortex (vmPFC), the hippocampus and the amygdala (Quirk and Mueller, 2008; Milad and Quirk, 2012). In addition to the passive extinction learning process, conditioned responses can be diminished by exerting instrumental control over the aversive event, thereby actively avoiding the threatening outcome (LeDoux and Gorman, 2001; Baratta *et al.*, 2007).

Received: 21 February 2020; **Revised:** 3 April 2020; **Accepted:** 2 June 2020

It has been shown almost half a century ago that instrumental control over aversive events may alleviate the deleterious effects of these events, whereas the lack of control over significant events may induce a state of learned helplessness that is characterized by serious cognitive, emotional as well as motivational deficits and may ultimately lead to psychopathology (Miller and Seligman, 1975). Animal studies provided initial insights into the neural mechanisms involved in exerting instrumental control over aversive events. Specifically, it has been shown that the medial prefrontal cortex detects control over aversive events and, as a result, inhibits activation of serotonergic neurons in the dorsal raphe nucleus, thereby preventing serotonergic signaling and learned helplessness (Amat *et al.*, 2005; Amat *et al.*, 2006; Amat *et al.*, 2008). However, although instrumental control over aversive events has been shown to reduce fear responses in humans (Hartley *et al.*, 2014), how this is implemented in the human brain and, in particular, to what extent the neural signature of instrumental control over aversive events deviates from the signature of passive extinction learning is not well understood.

The present experiment aimed at elucidating the neural underpinnings of learning how to actively control aversive events and to contrast these with those of passive extinction learning. Healthy participants performed an aversive learning task in the MRI scanner, during which one group of participants could learn how to avoid an electric shock (*controllability* group; $n = 21$), whereas participants in a second group were yoked to a participant in the controllability group, i.e. they were exposed to the exact same sequence of trials and shocks but without instrumental control over shock delivery (*yoked uncontrollability* group; $n = 21$). Whereas participants in the yoked uncontrollability group had no control over the shock, they received only a minimum number of shocks after their counterparts in the controllability group had learned the instrumental response, thus representing a yoked extinction process. We focused in particular on group differences after the time point of learning the instrumental response, which marked the onset of instrumental control in the controllability group but the onset of extinction in the yoked uncontrollability group. Participants in a third group did not receive any shocks and served as control group ($n = 22$). We predicted that instrumental control over aversive events would lead to a more pronounced decrease in the physiological fear response than passive extinction and that instrumental control over aversive events would be linked to medial prefrontal and striatal activity (LeDoux and Gorman, 2001; Cardinal *et al.*, 2002; Amat *et al.*, 2005; Amat *et al.*, 2006; Amat *et al.*, 2008), whereas fear extinction would be associated with the vmPFC (Quirk and Mueller, 2008; Milad and Quirk, 2012).

Materials and methods

Participants and experimental design

Seventy-five healthy, right-handed volunteers with normal or corrected-to-normal vision participated in this experiment. Exclusion criteria included past or present neurological or psychiatric disorders, medication intake, drug abuse and any MRI contraindications. The intended sample size was based on pilot testing and an a priori power calculation (Faul *et al.*, 2007), showing that this sample size enables the detection of a medium-sized behavioral effect of Cohen's $f = 0.25$ with a power of 0.95. Seven participants had to be excluded from all analyses because they received no electric shocks due to technical failure ($n = 2$), because they did not learn how to avoid shocks in the controllability condition (pre-defined criterion: more than 40 out of 50 possible shocks received and fewer than five consecutive shock-avoiding button presses; $n = 4$) or due to non-compliance with the instructions (lack of behavioral responding and excessive movement in the MRI, $n = 1$), resulting in a sample size of $n = 68$ for behavioral analyses (34 women; age: $M = 24.21$; s.d. = 3.39). Four additional participants were excluded for excessive head motion during the scan (>4.5 mm/degree in any direction), resulting in a final sample size of $n = 64$ for MRI analyses (33 women; age: $M = 24.28$; s.d. = 3.39). A *post hoc* power analysis confirmed that this final sample size was still sufficient to detect a medium-sized effect with a power of 0.94. All participants gave written informed consent before participation and received a monetary compensation of 30 EUR at the end of testing. The study protocol was approved by the local institutional review board (PV5120). Participants were randomly assigned to one of three different experimental groups: a controllability group ($n = 21$, 11 women), a yoked uncontrollability group ($n = 21$, 11 women) or a no-shock control group ($n = 22$, 11 women).

Aversive learning task

In order to assess how instrumental control over aversive events changes fear learning and its neural underpinnings, participants performed a learning task in which they received repeatedly moderate electric shocks (or not). Critically, while some participants could learn a behavioral response to avoid electric shocks (*controllability* group), others received an identical number of shocks at the exact same timings as their counterpart in the controllability group, but had no instrumental control over shock delivery (yoked *uncontrollability* group). In addition, a third group of participants received no shocks and served as control group (*no-shock control* group).

The task comprised 100 trials in total, 50 CS- trials and 50 CS+ trials for the controllability and uncontrollability group, while there were only CS- trials for the no-shock control group. Each trial started with a black fixation cross (6000–8000 ms) presented on a light gray background, followed by a conditioned stimulus consisting of a black frame (circle or square). Participants in the controllability and yoked uncontrollability group were instructed that one frame type (e.g. circle) could be followed by a brief (100 ms) single electric shock (unconditioned stimulus; UCS), thus representing the CS+, whereas the other stimulus (e.g. square) would never be followed by a shock and thus served as CS- (see Figure 1). Whether circle or square served as CS+ and CS-, respectively, was counterbalanced across participants. About 500–1000 ms after the onset of the CS, a black arrow pointing to the left or right appeared within the CS. Participants of all groups (including the no-shock control group) were instructed to press one out of four buttons whenever they saw an arrow on the screen. Participants in the controllability and yoked uncontrollability group received an additional instruction that they could avoid electric shocks after the CS+ by performing an instrumental action (i.e. a specific button press) but received no further information about the specific action. However, there was only an instrumental contingency between behavioral response and shock delivery in the controllability group. These instructions implied that participants in the uncontrollability group were deceived about the actual controllability of shocks, which may promote anger and frustration, but is thought to be an essential component of learned helplessness studies in humans

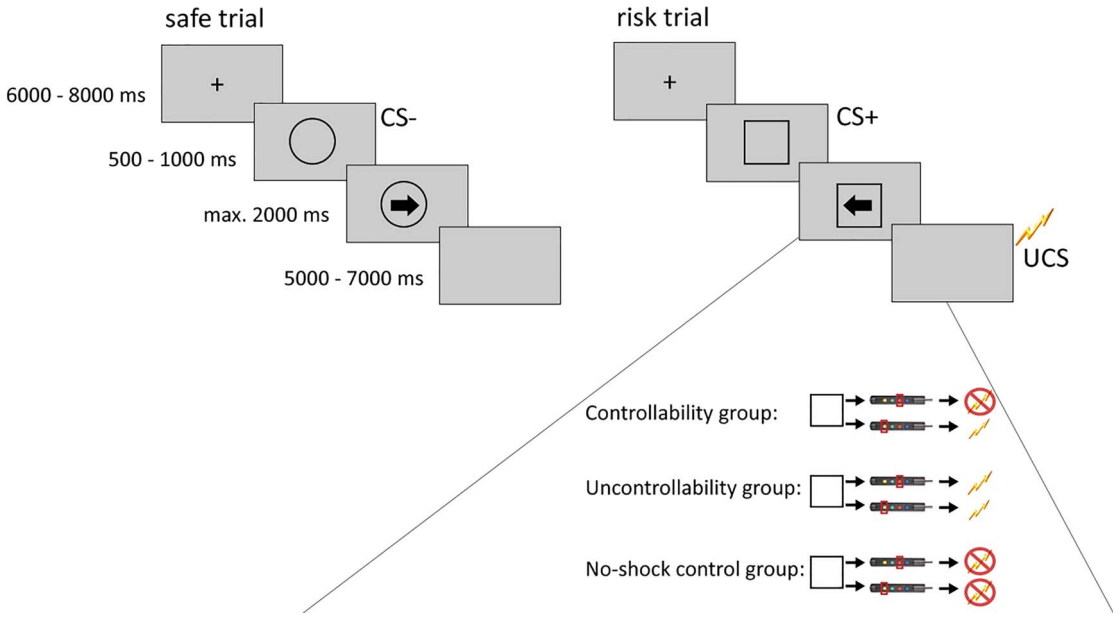

Fig. 1. Experimental design. Each trial started with the presentation of a fixation cross, followed by a circle or square, representing the CS- or CS+. After a brief interval, an arrow pointing to the left or right appeared within the CS and signaled to participants that they were required to press one out of four buttons on a four-button response box. Depending on the trial type and behavioral response, participants could receive a brief (100 ms) electric shock 5000–7000 ms after CS offset. In CS+ trials, participants in the controllability group could avoid the majority of electric shocks (90% of CS+ trials) by pressing the third button from the left on the response button box within a maximum response window of 2500 ms, regardless of whether the arrow pointed to the left or right. Accordingly, participants in the controllability group received an electric shock in CS+ trials when they pressed one of the three other buttons, gave no response within the response window, or in CS+ trials in which shock delivery could not be avoided (10% of CS+ trials). As participants in the uncontrollability group were yoked to a participant in the controllability group, they were exposed to the exact same sequence of trial types and shocks as their counterpart in the controllability group. Thus, they received an electric shock in CS+ trials in which their yoked counterpart had received a shock, irrespective of the behavioral response. The no-shock control group did not receive any shocks during the experiment.

(Mikulincer, 1994). While the first three CS+ were always followed by a shock to ensure CS+/CS- differentiation from the beginning of the task on, participants in the controllability group could avoid shocks in 90% of the remaining CS+ trials by pressing the third button on a four-button response box within a maximum response window of 2000 ms after arrow onset, irrespective of the direction the arrow pointed to. Consequently, participants in the controllability group were exposed to a shock after CS+ presentation if they gave a wrong or no behavioral response; additionally, they received a shock in 10% of the CS+ trials regardless of their actual response (see below). In contrast, participants in the yoked uncontrollability group were unable to exert instrumental control over shock delivery at any point during the experiment. They were yoked to a participant in the controllability group, meaning that exactly the same sequence of trials and shocks from a participant of the controllability group was replayed for their yoked participant in the uncontrollability group. Hence, while participants in the controllability and uncontrollability groups experienced an equal number of shocks following the CS+ at the exact same timings, there was no contingency between behavioral response and shock delivery in the uncontrollability group. Notably, the exclusion of participants as described above did only minimally affect our yoking procedure, as the majority of excluded participants were already replaced during data collection. Specifically, a yoked pairing did not exist for one participant in the controllability and one participant in the uncontrollability group in the fMRI analysis due to movement and one participant in the skin conductance response (SCR) analysis due being a non-responder.

Whereas participants in the controllability groups could avoid shocks in 90% of the trials, in 10% of the trials they received a shock irrespective of their response. The rationale behind this procedure was to prevent a clear-cut end of electric shocks after participants in the controllability group had learned the correct shock-avoiding response and, as a result, a potential illusory control in the uncontrollability group. To avoid that this would affect the perceived controllability, participants in the controllability group were instructed that they could 'significantly decrease the risk of receiving a shock when performing the correct response', whereas participants in the uncontrollability group were instructed that they could 'reliably avoid shocks when performing the correct response'.

The trial order was pseudo-randomized to avoid that more than three trials of the same type (CS+ *vs* CS-) occurred in a row. CS and arrow remained on the screen until the participant made a response or until a maximum response time interval of 2000 ms after arrow onset and the time interval between CS+ offset and shock delivery was 5000–7000 ms (randomly jittered).

Experimental procedure

Participants completed questionnaires and received task instructions before we collected baseline saliva samples and measured blood pressure and pulse. We attached disposable electrodes to participants left palms for skin conductance recordings. In the controllability and uncontrollability group, we also placed disposable electrodes to participants' right lower legs for electric stimulation before we selected an individual shock intensity that was rated to be 'unpleasant, but not yet painful'. Then participants performed an unrelated working memory task (about 15 min) before they underwent the manipulation of

controllability (about 25 min) in the MRI scanner. At the end of the experiment, participants completed a rating questionnaire and participants in the uncontrollability group were debriefed that they had received deceptive instructions regarding the controllability of electric shocks.

Behavioral analysis

We refer to behavioral responses (i.e. the button presses) as 'shock-avoiding' if they avoided shocks in 90% of CS+ trials in the controllability group, whereas all other responses are referred to as 'not shock-avoiding'.

In order to disentangle the physiological and neuronal responses during fear acquisition from those during instrumental control (controllability group) or extinction learning (yoked uncontrollability group), we further identified an individual time point of learning for each participant in the controllability group. Based on pilot data (supplementary material), we defined the first out of the first five consecutive shock-avoiding button presses in CS+ trials as time point of learning. Both CS+ and CS- trials of the controllability participant and the yoked uncontrollability and no-shock control participant that had been randomly assigned to a triplet during testing [referred a 'triadic design' (Maier and Seligman, 1976)] were split into a before learning and after learning phase according to this time point. We could not identify a time point of learning for one participant in the controllability group (fewer than five consecutive shock-avoiding responses) and, thus, had to exclude this participant and the corresponding yoked participant ($n = 2$) from all analyses involving the factor learning.

Splitting trial types into a before and after learning phase enabled us to investigate differences in physiological and neuronal processes related to fear acquisition (before learning) and processes occurring after the time point of learning, marking detection of instrumental control over the UCS in the controllability group, i.e. the onset of instrumental control, and the onset of an extinction process in the yoked uncontrollability group.

Acquisition and analysis of SCRs

Skin conductance data were acquired with a Biopac-MP-160 sampling module (BIOPAC Systems, Goleta, USA) and analyzed using Ledalab (Version 3.4.9; Benedek and Kaernbach, 2010). SCR data were downsampled to a resolution of 10 Hz, filtered with a 5 Hz low-pass filter and then analyzed using continuous decomposition analysis as implemented in Ledalab.

Acquisition and analysis of fMRI data

Images were collected using a 3 T Siemens Prisma Scanner with a 64-channel head coil. The neuroimaging data were analyzed using SPM12 (the Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/software/spm12/), running under MATLAB R14a (MathWorks Inc., Natick MA, USA). Our analysis included a standard pre-processing procedure (spatial realignment, coregistration, normalization and smoothing) and general linear modeling. We used region of interest (ROI) analyses; to this end, we selected a priori ROIs based on previous literature on neuronal underpinnings of controllability and created a combined mask containing these ROIs using Marina (http://www.bion.de/eng/MARINA.php) to correct for α-error accumulation associated with testing multiple ROIs. We additionally used a psychophysiological interaction (PPI) analysis as implemented in SPM12 to assess controllability-dependent connectivity changes between the vmPFC and the amygdala as the vmPFC–amygdala cross talk is assumed to play a critical role in the reduction of fear (Milad and Quirk, 2002; Bouton *et al.*, 2006; Adhikari *et al.*, 2015).

A more detailed description of the experimental procedure, behavioral analyses, skin conductance and imaging analyses is provided in the supplemental material.

Results

Successful acquisition of the shock-avoiding instrumental response in the controllability group

Participants in the controllability group learned the instrumental response required to avoid electric shocks after the CS+ very well across the task ($F(49\ 1029) = 5.505$, $P < 0.001$, $\eta^2 = 0.208$). The number of shock-avoiding responses in the controllability group in CS+ trials differed further significantly from those in the yoked uncontrollability group ($t(43) = 9.308$, $P < 0.001$, $d = 2.839$) and no-shock control group ($t(32.320) = 14.954$, $P < 0.001$, $d = 5.261$; GROUP × TRIAL interaction: $F(98\ 1285.700) = 2.991$, $P < 0.001$, $\eta^2 = 0.085$; see Figure 2). The total number of shocks participants received during the experiment differed minimally due to the exclusion of participants after data collection [due to movement (fMRI analysis): controllability group, $n = 1$; uncontrollability group, $n = 1$; due to non-responsivity in SCR, controllability group, $n = 1$]. These differences, however, were very small and not statistically significant [number of shocks for fMRI analysis, $M_{CON} = 19.857$ (s.d. = 8.101), $M_{UNCON} = 20.095$ (s.d. = 8.178), $t(40) = 0.095$, $P = 0.925$; and for SCR analysis, $M_{CON} = 20.524$ (s.d. = 7.910), $M_{UNCON} = 20.091$ (s.d. = 7.982), $t(41) = 0.179$, $P = 0.859$]. We further determined individual time points of learning the instrumental response for each participant in the controllability group, showing that participants needed on average 16.81 CS+ trials (s.d. = 8.43; range: 6–37 trials to criterion) to acquire the instrumental response directed at avoiding shocks after the CS+.

Detection of instrumental control over the UCS reduces threat-related SCRs

Next, we investigated how our experimental manipulation affected psychophysiological responses to CS+ and CS- onset. We first investigated whether there was a general conditioning effect indicated by larger SCR in response to the CS+ compared to the CS- in the two experimental groups that had received shocks (controllability and yoked uncontrollability groups). As expected, participants in these groups showed a robust conditioning effect, reflected in larger SCRs to the CS+ compared to the CS- ($t(41) = 3.045$, $P = 0.004$, $d = 0.951$). In contrast, SCRs to both CS types did not differ in the no-shock control group ($t(19) = 0.151$, $P = 0.881$, $d = 0.069$).

As we were primarily interested in changes in responding related to the detection of instrumental control, in contrast to (yoked) extinction, we next included the time point of (instrumental) learning into our analysis, which marked the onset of instrumental control in the controllability group and the onset of extinction in the yoked uncontrollability group. To this end, we entered SCRs to CS+ and CS- onset before and after the time point of learning into a mixed model analysis of variance (ANOVA) involving the factors trial type (CS+ *vs* CS-), learning (before *vs* after learning) and group (controllability *vs*

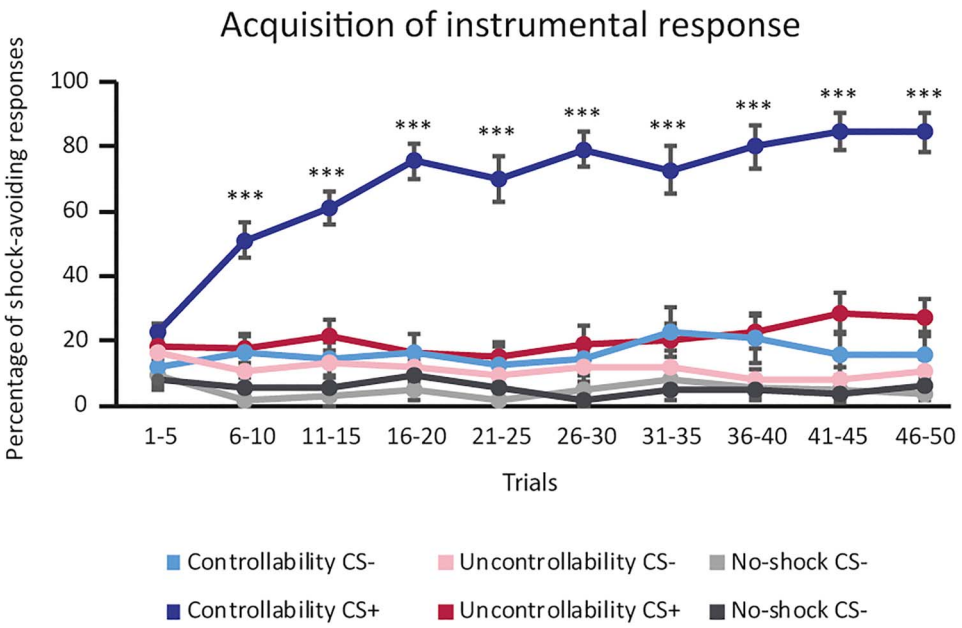

Fig. 2. Behavioral results. Percentage of shock-avoiding responses during the course of the experiment for experimental groups and trial types. Behavioral responses directed at avoiding electric shocks increased across 5-trial blocks in the controllability group, indicating successful instrumental learning, whereas there was no such increase in the yoked uncontrollability and no-shock control groups. Error bars represent standard errors of the mean. *** $P < 0.001$ (uncorrected).

yoked uncontrollability *vs* no-shock control group). This analysis showed that changes in SCR from before to after learning differed significantly between groups (GROUP × TRIAL TYPE × LEARNING interaction, $F(2.59) = 5.088$, $P = 0.009$, $\eta^2 = 0.147$). Following up on this interaction, we observed that both the controllability and yoked uncontrollability group tended to show larger SCRs to CS+ than to the CS- before the time point of (instrumental) learning (controllability group, $t(19) = 2.311$, $P_{corr} = 0.096$, $d = 1.060$; yoked uncontrollability group, $t(21) = 2.520$, $P_{corr} = 0.060$, $d = 1.100$; without group differences, $F(1.40) = 1.069$, $P = 0.307$, $\eta^2 = 0.026$; as opposed to the no-shock control group ($t(19) = 0.105$, $P_{corr} > 0.999$, $d = 0.048$). However, striking differences between the controllability and uncontrollability groups were observed after the time point of (instrumental) learning: in the controllability group, there was no difference in SCRs to CS+ and CS- anymore after learning the instrumental shock-avoiding response ($t(19) = 0.009$, $P_{corr} > 0.999$, $d = 0.004$), same as in the no-shock control group ($t(19) = 0.091$, $P_{corr} > 0.999$, $d = 0.042$). Participants in the yoked uncontrollability group, in turn, still tended to show significantly larger SCRs to the CS+ than to the CS- after the time point of (instrumental) learning ($t(21) = 2.573$, $p_{corr} = 0.054$, $d = 1.123$; see Figure 3). While both instrumental control and extinction diminished SCRs to CS+ presentation from pre- to post-learning (controllability group, $t(19) = 4.179$, $P_{corr} = 0.003$, $d = 1.917$; yoked uncontrollability group, $t(21) = 5.883$, $P_{corr} < 0.003$, $d = 2.568$), these results indicate that instrumental control resulted in a stronger reduction of threat responses as shown by a lack of differential SCR between CS+ and CS- after learning in the controllability group (between-group difference for differential CS+ minus CS- SCRs in controllability *vs* uncontrollability, $t(40) = 2.233$, $P = 0.031$, $d = 0.706$).

Brain regions associated with processing threat and safety

We next identified brain regions that were involved in processing threat (CS+) and safety (CS-) across the controllability and yoked uncontrollability groups. The overall contrast (CS+ > CS-) yielded clusters in striatal regions (right: [14 6–6], $P_{svc} = 0.006$, FWE-corrected, $T = 5.29$, $k = 66$; left: [−10 6–4], $P_{svc} = 0.054$, FWE-corrected, $T = 4.45$, $k = 42$), which is in line with previous studies, indicating that the striatum shows robust responses to threat-predictive cues (Fullana *et al.*, 2016). The reverse contrast (CS- onset >CS+ onset) revealed one cluster involving bilateral vmPFC ([6 56–10], $P_{svc} < 0.001$, FWE-corrected, $T = 6.85$, $k = 857$) and a cluster in the left putamen, extending to the insula, pre- and post-central and temporal gyri ([−30–10 14], $P_{svc} = 0.002$, FWE-corrected, $T = 5.65$, $k = 49$). As the no-shock control group did not experience threat/safety, we performed a separate contrast for this group and observed no significant clusters, indicating that neural responses did not differ between the two CS.

Different neural mechanisms underlie instrumental control and extinction

We then tackled our primary research question, i.e. how instrumental control over threatening outcomes and extinction, respectively, affected the processing of the CS+ and the CS-. To this end, we ran a $3 \times 2 \times 2$ full factorial model including the factors group (controllability *vs* yoked uncontrollability group *vs* no-shock control), trial type (CS+ *vs* CS-) and learning (before *vs* after learning). This analysis yielded significant GROUP × TRIAL TYPE × LEARNING interactions in two clusters, one including the left caudate, putamen and thalamus ([−10 6 10], $P_{svc} = 0.007$, FWE-corrected, $F = 13.62$, $k = 178$) and one including the left caudate and ACC ([−16 26–4], $P_{svc} = 0.011$, FWE-corrected, $F = 12.94$, $k = 50$). We further observed one cluster at trend level involving the right caudate ([12 6 6], $P_{svc} = 0.053$, FWE-corrected, $F = 10.99$, $k = 152$).

In order to pursue this interaction, we performed separate TRIAL TYPE × LEARNING full factorial models for each group alone. In the controllability group, this analysis yielded significant two-way interactions in bilateral clusters in the

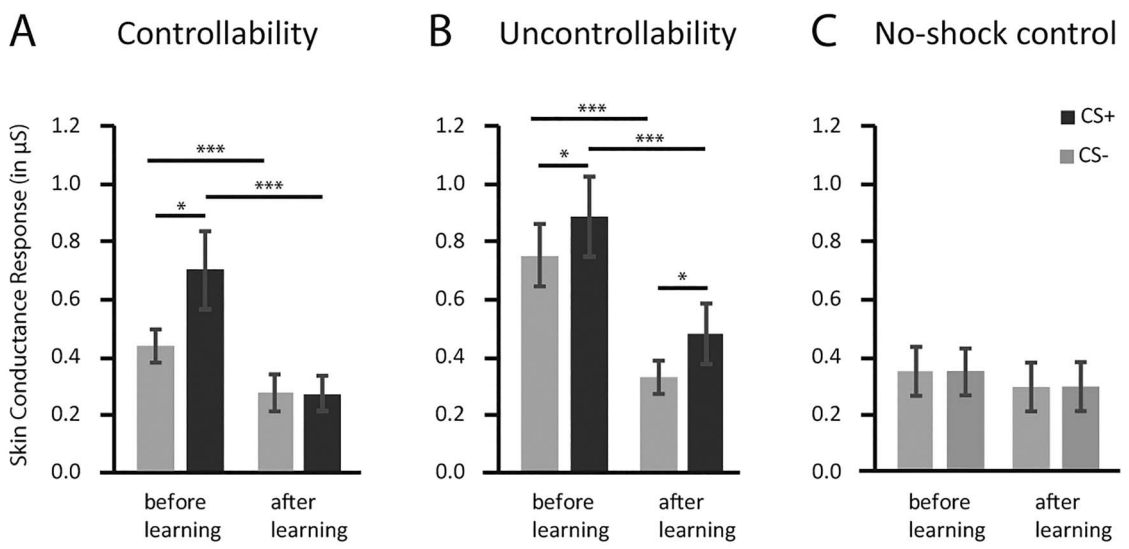

Fig. 3. Instrumental control leads to a more pronounced reduction of SCRs than yoked extinction. Mean SCRs to the different CS types before and after the time point of successful instrumental learning in the **A** controllability group, **B** uncontrollability group and **C** no-shock control group. Before learning, SCRs to the CS+ were significantly higher than to the CS- in both the controllability and uncontrollability group. Both the controllability and yoked uncontrollability group showed significantly reduced SCRs to the CS+ after learning as compared to the CS+ before learning; however, this reduction of SCRs to the CS+ after learning was more pronounced in the controllability group. SCRs to the CS+ did not differ from SCRs to the CS- after learning in the controllability group, whereas SCRs to the CS+ remained significantly higher than SCRs to the CS- in the yoked uncontrollability group, indicating that instrumental control diminishes psychophysiological fear to a larger extent than yoked extinction. Error bars represent standard errors of the mean. $^{*}P < 0.05$ (uncorrected), $^{***}P < 0.001$ (uncorrected).

caudate and putamen (left: [−12 6 4], $P_{SVC} = 0.005$, FWE-corrected, $F = 27.01$, $k = 158$; right: [12 8 4], $P_{SVC} = 0.018$, FWE-corrected, $F = 23.17$, $k = 489$). Follow-up tests revealed that all of these interactions were driven by decreased activation to the CS+ after learning >CS- after learning as compared to the CS+ before learning >CS- before learning (left: [−12 4 6], $P_{SVC} = 0.015$, FWE-corrected, $T = 5.10$, $k = 119$; right: [20 8 10], $P_{SVC} = 0.016$, FWE-corrected, $T = 5.09$, $k = 415$).

In contrast, the 2 × 2 model for the yoked uncontrollability group yielded a significant interaction in the left vmPFC ([−4 30–20], $P_{SVC} = 0.033$, FWE-corrected, $F = 20.90$, $k = 42$). A follow-up *t*-test showed that the effects observed in the vmPFC were driven by significantly increased activation to the CS+ after learning >CS- after learning as compared to the CS+ before learning >CS- before learning ([−6 28–20], $P_{SVC} = 0.047$, FWE-corrected, $T = 4.56$, $k = 28$).

A 2 × 2 model for the no-shock control group did not yield any significant interactions (all $P_{SVC} > 0.522$, FWE-corrected).

Together, this pattern of results indicates that different brain areas are involved in instrumental control and extinction processes, with decreased activation to threat cues in bilateral striatal regions in the controllability group and increased activation to threat cues in the left vmPFC in the yoked uncontrollability group (see Figure 4).

Instrumental control over the UCS increases functional connectivity of the vmPFC with the amygdala

To investigate whether instrumental control over aversive stimuli alters the crosstalk between the vmPFC and the amygdala, two brain regions known to be critically involved in the reduction of fear (Milad and Quirk, 2002; Bouton *et al.*, 2006; Adhikari *et al.*, 2015), we performed a functional connectivity analysis using a PPI. We chose the right vmPFC ([8 32–10]) as seed and created a sphere with a radius of 6 mm around the peak coordinate. This PPI analysis revealed that the controllability group showed increased crosstalk between the right vmPFC and left amygdala ([−22–2 -26], $P_{SVC} = 0.046$, FWE-corrected, $T = 3.05$, $k = 10$) relative to the yoked uncontrollability group (see Figure 5). We further obtained evidence for a basolateral location of the obtained cluster using separate masks for the basolateral, centromedial and superficial amygdala taken from the Juelich Atlas (Amunts *et al.*, 2005; see supplemental material).

Discussion

Fear conditioning is assumed to be a key process in the pathogenesis of anxiety disorders (Graham and Milad, 2011; VanElzakker *et al.*, 2014; Duits *et al.*, 2015). Accordingly, mechanisms that may reduce fear are of great interest. In the present experiment, we directly contrasted the psychophysiological and neuronal signatures of passive extinction learning, on which most therapeutic interventions for anxiety disorders rely, with those of instrumental control over aversive events. We show that while both (yoked) extinction and instrumental control decreased SCRs as established indicator of conditioned fear, this decrease was significantly more pronounced for instrumental control. Moreover, our data show that extinction learning and instrumental control are subserved by distinct neural mechanisms. Whereas passive extinction learning was associated with increased vmPFC activity, instrumental control was linked to a decrease in striatal activity.

While initial fear acquisition was comparable in the controllability and yoked uncontrollability groups, the subsequent decrease in SCRs to threat cues was significantly stronger after acquiring instrumental control over aversive events than during passive extinction. Instrumental control rapidly eliminated the physiological fear response, whereas a residual fear response remained after passive extinction. The remaining SCR to the CS+ may be owing to the infrequent US presentations during the yoked extinction process. These infrequent US presentations, however, were not sufficient to provoke threat-related SCRs in

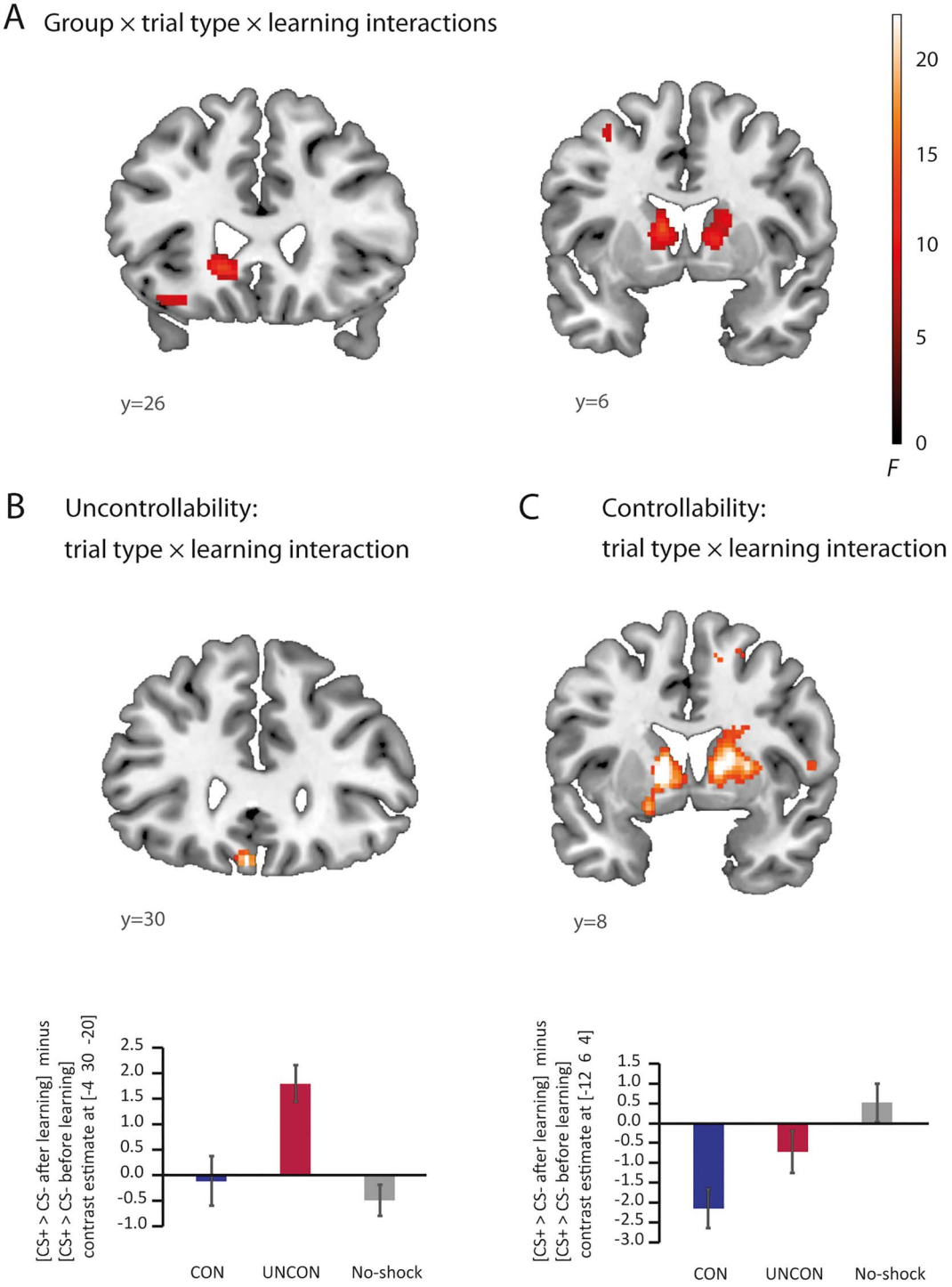

Fig. 4. Different neural mechanisms for instrumental control and yoked extinction. **A** Results of a univariate GROUP × TRIAL TYPE × LEARNING interaction revealed significant clusters in the left ventromedial prefrontal cortex (vmPFC) and bilateral striatum. **B** For the uncontrollability group, a follow-up analysis revealed a TRIAL TYPE × LEARNING interaction in the vmPFC, which was driven by increased vmPFC activation after learning as compared to before learning. **C** In the controllability group, a follow-up analysis showed a TRIAL TYPE × LEARNING interaction in the bilateral striatum, which was driven by decreased striatal activation after learning as opposed to before learning. For visualization purposes, all clusters are displayed at a threshold of $P < 0.001$, uncorrected. Please note that contrast estimates depicted for visualization purposes are extracted from peak voxels of clusters obtained in the ROI analysis and thus do not represent independent analyses.

individuals who had extensive instrumental control over shock delivery. It has been shown before that control over aversive events significantly alters affective processing, in a way that protects organisms against the negative responses following aversive events (Abramson *et al.*, 1978). Even after exposure to uncontrollable aversive events, learning to exert control over the aversive event can alleviate the negative consequences following from the uncontrollability experience (LeDoux and Gorman, 2001; Baratta *et al.*, 2007; Hartley *et al.*, 2014). In contrast to earlier studies (Boeke *et al.*, 2017; Hartley *et al.*, 2019), both the

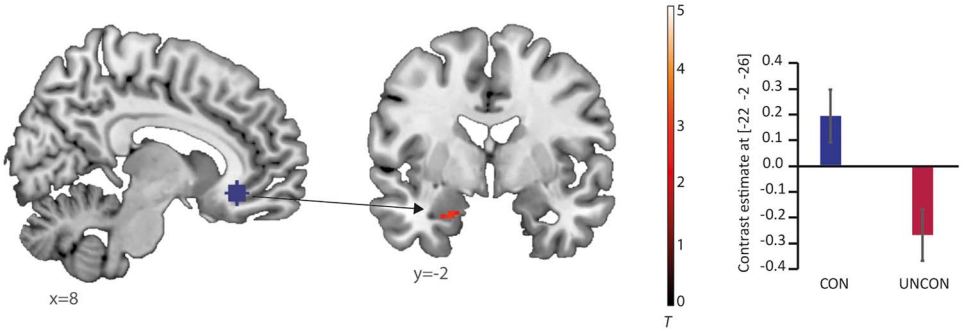

Fig. 5. Between-group differences in functional connectivity between the ventromedial prefrontal cortex (vmPFC) and the amygdala. Visualization of between-group differences in functional connectivity between the right vmPFC and the left amygdala for the contrast [PPI 'CS+' > PPI 'CS-']. The blue area represents a 6 mm sphere around peak level coordinate in the right vmPFC from a univariate GROUP × TRIAL TYPE interaction ([8 32–10]); the red area represents between-group differences in cross talk between the vmPFC and the left amygdala, for visualization purposes displayed at $P < 0.001$, uncorrected. The controllability group showed increased cross talk between the vmPFC and the left amygdala when compared to the uncontrollability group. Please note that contrast estimates depicted for visualization purposes are extracted from peak voxels of clusters obtained in the ROI analysis and thus do not represent independent analyses.

controllability and uncontrollability groups received in our study the instruction that they could learn to perform a behavioral response that would avoid the aversive outcome, suggesting that the results reported here are not driven by differential instructions affecting top-down processing.

In addition to a more pronounced reduction of fear responses following the detection of instrumental control over aversive outcomes, our results further suggest that instrumental control and passive extinction processes can be disentangled at the level of the blood-oxygen-level-dependent (BOLD) signal. In the controllability group that could exert instrumental control over aversive events, the reduction of threat-related psychophysiological arousal was paralleled by decreased BOLD signal in the striatum. The striatum has been implicated in the anticipation of aversive stimuli and threat processing (Jensen *et al.*, 2003; Fullana *et al.*, 2016); thus the decrease of striatal activation in response to threat cues after successful instrumental response learning indicates some form of safety learning that the initial threat cue is no longer threatening. Passive extinction learning, in turn, was associated with increased BOLD signal in the vmPFC. Although this association was not present in our full model testing for a three-way interaction and should therefore be interpreted with caution, this corresponds to previous research that has established a key role of the vmPFC in fear extinction through descending projections to brain regions involved in fear expression/inhibition, such as the amygdala, the brainstem and the hypothalamus (Quirk and Mueller, 2008). Our findings are further well in line with an earlier study on the subject of control that reported reduced vmPFC activation in response to controllable *vs* uncontrollable aversive events (Wood *et al.*, 2015). However, another study observed increased vmPFC activation during controllable as opposed to uncontrollable trials (Kerr *et al.*, 2012). These divergent findings may stem from differences in study designs: while the first study used a yoked design similar to our design, the second study employed a within-subject design that did thus not only involve control and a lack of control but also loss of previously experienced control. Although lack of control and loss of control may appear to describe very similar conditions, early research in the field of learned helplessness found this to be an important distinction (Maier and Seligman, 1976). Hence, it is crucial to consider experimental designs when discussing effects of control over aversive outcomes.

Beyond changes in the activity of brain areas highly relevant in the context of fear learning, extinction and control, we further observed differences in neural coupling between the vmPFC and the amygdala in yoked extinction and instrumental control. We observed increased crosstalk between the vmPFC and the (basal) amygdala in the group that had instrumental control as compared to the yoked extinction group. In rodents, it has been demonstrated that the basal nucleus of the amygdala is required for successful avoidance learning (Mogenson *et al.*, 1980; Robbins *et al.*, 1989), but not for passive fear responses (Amorapanth *et al.*, 2000), and it has further been shown that projections from the vmPFC to the basal nucleus of the amygdala are crucially involved in top-down suppression of fear responses (Milad and Quirk, 2002; Bouton *et al.*, 2006; Adhikari *et al.*, 2015). Similarly, there is evidence in humans indicating that the coupling between the vmPFC and the amygdala is associated with successful instrumental learning how to avoid an aversive outcome (Collins *et al.*, 2014). Thus, the increased crosstalk between the vmPFC and the basal amygdala in the group that had instrumental control over aversive outcomes that we observed here may reflect a mechanism that is specific to successful instrumental avoidance of aversive events and might further reflect top-down regulation of fear responses underlying the reduction of psychophysiological fear responses.

Given the link between experiences of uncontrollability over significant life events and the pathogenesis of mental disorders, it is important to identify mechanisms that are involved in the detection of control in order to detect disturbances that may point to clinical and pre-clinical conditions. While earlier research either instructed participants explicitly how to avoid aversive events in controllable trials (Wood *et al.*, 2015) or employed responses that could be learned very quickly (Boeke *et al.*, 2017), which meant that control was present from the beginning of the experiment, we used here a paradigm that enabled us to individually define a pre- and post-learning phase for participants in the controllability group and to identify changes in neural and physiological responses after the onset of control. This approach provides new mechanistic insight into the detection of control in the aversive domain and may bear the potential to serve as clinical predictor or target for interventions.

Finally, we acknowledge that the yoked extinction procedure used here differs from standard extinction protocols that do not involve UCS presentations during the extinction phase. Although our yoked extinction phase included only very few shocks, it seems possible that this led to higher shock expectations and a delayed extinction process as compared to standard extinction.

In sum, we show here critical differences in the psychophysiological effectiveness and neural underpinnings of two fear reduction mechanisms, passive extinction learning and instrumental control over threatening events. Whereas passive extinction resulted in a decrease of fear-related SCRs, this decrease was significantly more pronounced after acquiring instrumental control over aversive events. Extinction learning further involved the vmPFC, a key region in the fear extinction network (Quirk and Mueller, 2008), whereas instrumental control was primarily associated with reduced striatal responses and increased coupling between the vmPFC and the amygdala. Together, these findings shed light on the neural signature of instrumental control as a potent modulator of conditioned fear and might have important implications for the treatment of fear-related disorders.

Supplementary material

Supplementary material is available at *SCAN* online.

Acknowledgements

The authors thank Laura Clausen, Kim Löwisch, Jennifer Maurer and Lena Piel for their assistance during data collection and Christian Büchel for his helpful comments on the experimental design. The authors further acknowledge the help of Dhana Bräuer, Daniel Kutzner, Christoph Schäde, Jan Sedlacik, Anne Tiefert and Anja Turlach during fMRI scanning.

Conflict of interest

The authors declare no competing financial interests.

References

Abramson, L.Y., Seligman, M.E., Teasdale, J.D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, **87**(1), 49–74.

Adhikari, A., Lerner, T.N., Finkelstein, J., *et al.* (2015). Basomedial amygdala mediates top-down control of anxiety and fear. *Nature*, **527**(7577), 179–85.

Amat, J., Baratta, M.V., Paul, E., Bland, S.T., Watkins, L.R., Maier, S.F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience*, **8**(3), 365–71.

Amat, J., Paul, E., Zarza, C., Watkins, L.R., Maier, S.F. (2006). Previous experience with behavioral control over stress blocks the behavioral and dorsal raphe nucleus activating effects of later uncontrollable stress: role of the ventral medial prefrontal cortex. *The Journal of Neuroscience*, **26**(51), 13264–72.

Amat, J., Paul, E., Watkins, L.R., Maier, S.F. (2008). Activation of the ventral medial prefrontal cortex during an uncontrollable stressor reproduces both the immediate and long-term protective effects of behavioral control. *Neuroscience*, **154**(4), 1178–86.

Amorapanth, P., LeDoux, J.E., Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nature Neuroscience*, **3**(1), 74–9.

Amunts, K., Kedo, O., Kindler, M., *et al.* (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl)*, **210**(5–6), 343–52.

Baratta, M.V., Christianson, J.P., Gomez, D.M., *et al.* (2007). Controllable versus uncontrollable stressors bi-directionally modulate conditioned but not innate fear. *Neuroscience*, **146**(4), 1495–503.

Benedek, M., Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, **190**(1), 80–91.

Boeke, E.A., Moscarello, J.M., LeDoux, J.E., Phelps, E.A., Hartley, C.A. (2017). Active avoidance: neural mechanisms and attenuation of Pavlovian conditioned responding. *The Journal of Neuroscience*, **37**(18), 4808–18.

Bouton, M.E., Westbrook, R.F., Corcoran, K.A., Maren, S. (2006). Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biological Psychiatry*, **60**(4), 352–60.

Calhoon, G.G., Tye, K.M. (2015). Resolving the neural circuits of anxiety. *Nature Neuroscience*, **18**(10), 1394–404.

Cardinal, R.N., Parkinson, J.A., Hall, J., Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, **26**(3), 321–52.

Collins, K.A., Mendelsohn, A., Cain, C.K., Schiller, D. (2014). Taking action in the face of threat: neural synchronization predicts adaptive coping. *The Journal of Neuroscience*, **34**(44), 14733–8.

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

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**(2), 175–91.

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., *et al.* (2016). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, **21**(4), 500–8.

Graham, B.M., Milad, M.R. (2011). The study of fear extinction: implications for anxiety disorders. *The American Journal of Psychiatry*, **168**(12), 1255–65.

Hartley, C.A., Gorun, A., Reddan, M.C., Ramirez, F., Phelps, E.A. (2014). Stressor controllability modulates fear extinction in humans. *Neurobiology of Learning and Memory*, **113**, 149–56.

Hartley, C.A., Coelho, C.A.O., Boeke, E., Ramirez, F., Phelps, E.A. (2019). Individual differences in blink rate modulate the effect of instrumental control on subsequent Pavlovian responding. *Psychopharmacology*, **236**(1), 87–97.

Jensen, J., McIntosh, A.R., Crawley, A.P., Mikulis, D.J., Remington, G., Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron*, **40**(6), 1251–7.

Kerr, D.L., McLaren, D.G., Mathy, R.M., Nitschke, J.B. (2012). Controllability modulates the anticipatory response in the human ventromedial prefrontal cortex. *Frontiers in Psychology*, **3**, 557.

LeDoux, J.E., Gorman, J.M. (2001). A call to action: overcoming anxiety through active coping. *The American Journal of Psychiatry*, **158**(12), 1953–5.

Maier, S.F., Seligman, M.E. (1976). Learned helplessness: theory and evidence. *Journal of Experimental Psychology: General*, **105**(1), 3–46.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual Review of Neuroscience*, **24**, 897–931.

Maren, S., Quirk, G.J. (2004). Neuronal signalling of fear memory. *Nature Reviews. Neuroscience*, **5**(11), 844–52.

Mikulincer, M. (1994). *Human Learned Helplessness: A Coping Perspective*, New York: Plenum.

Milad, M.R., Quirk, G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, **420**(6911), 70–4.

Milad, M.R., Quirk, G.J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annual Review of Psychology*, **63**, 129–51.

Milad, M.R., Pitman, R.K., Ellis, C.B., *et al.* (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biological Psychiatry*, **66**(12), 1075–82.

Miller, W.R., Seligman, M.E. (1975). Depression and learned helplessness in man. *Journal of Abnormal Psychology*, **84**(3), 228–38.

Mogenson, G.J., Jones, D.L., Yim, C.Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Progress in Neurobiology*, **14**(2–3), 69–97.

Pavlov, P.I. (2010). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, **17**(3), 136–41.

Quirk, G.J., Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, **33**(1), 56–72.

Rauch, S.L., Shin, L.M., Phelps, E.A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research–past, present, and future. *Biological Psychiatry*, **60**(4), 376–82.

Robbins, T.W., Cador, M., Taylor, J.R., Everitt, B.J. (1989). Limbic-striatal interactions in reward-related processes. *Neuroscience and Biobehavioral Reviews*, **13**(2–3), 155–62.

Rosen, J.B., Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review*, **105**(2), 325–50.

VanElzakker, M.B., Dahlgren, M.K., Davis, F.C., Dubois, S., Shin, L.M. (2014). From Pavlov to PTSD: the extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiology of Learning and Memory*, **113**, 3–18.

Wood, K.H., Wheelock, M.D., Shumen, J.R., Bowen, K.H., Ver Hoef, L.W., Knight, D.C. (2015). Controllability modulates the neural response to predictable but not unpredictable threat in humans. *NeuroImage*, **119**, 371–81.