




ARTICLE

Neural responses during extinction learning predict exposure therapy outcome in phobia: results from a randomized-controlled trial

Iris Lange¹, Liesbet Goossens¹, Stijn Michiels¹ , Jindra Bakker¹, Bram Vervliet², Machteld Marcelis^{1,3}, Marieke Wichers⁴, Jim van Os^{1,5,6}, Therese van Amelsvoort¹ and Koen Schruers^{1,7}

Extinction learning is assumed to represent a core mechanism underlying exposure therapy. Empirical evaluations of this assumption, however, are largely lacking. The current study investigated whether neural activations and self-report outcomes during extinction learning and extinction recall could specifically predict exposure therapy response in specific phobia. In this double-blind randomized controlled trial, individuals with spider phobia ($N = 45$; female/male = 41/4) were on group basis randomly allocated to exposure therapy ($n = 25$; female/male = 24/1) or progressive muscle relaxation (PMR; $n = 20$; female/male = 17/3). Intervention effects were measured with the Fears of Spiders questionnaire. Participants also underwent a three-day fear conditioning, extinction learning, and extinction recall paradigm during functional magnetic resonance imaging at baseline. Extinction outcomes were self-reported fear and threat expectancy, and neural responses during conditioned stimulus processing and during extinction-related prediction errors (US omissions) in regions of interest (ventromedial prefrontal cortex (vmPFC) and nucleus accumbens). Results showed that exposure therapy resulted in stronger symptom reductions than PMR (Cohen's $d = 0.90$). Exposure therapy response was specifically predicted by prediction-error related vmPFC activation during early extinction. There were also indications vmPFC activations during conditioned safety stimulus processing at early extinction predicted therapy outcome. Neural activations during extinction recall and self-report data did however not predict therapy outcome. These findings indicate that exposure therapy may rely on neural extinction learning processes. Prediction errors are thought to drive the extinction learning process, and prediction error-related vmPFC activation specifically predicted therapy outcome. The extent to which vmPFC processes safety signals may additionally be predictive of exposure therapy response, but the specificity is less clear.

Neuropsychopharmacology (2020) 45:534–541; <https://doi.org/10.1038/s41386-019-0467-8>

INTRODUCTION

Exposure therapy is regarded as one of the most effective treatments for anxiety disorders, with several studies demonstrating large effect sizes that can be maintained over long follow-up periods [1–3]. Yet, non-response and relapse occurs in about half of the patients [4]. To improve current therapeutic options, it is critical to increase existing insight into the mechanisms underlying exposure therapy [5]. Extinction learning has been regarded as the core principle underlying exposure therapy [6]. Over many years, preclinical and human experimental research has employed extinction-learning procedures in the laboratory to model exposure therapy. The validity of extinction learning paradigms as a model of exposure therapy has however not been thoroughly established [7]. More specifically, tests of the predictive validity, i.e., whether inter-individual differences in extinction learning can predict exposure therapy success, are still largely missing [7].

To investigate extinction learning in the laboratory, a fear-conditioning phase is first presented during which the association between a conditioned stimulus (CS+) and an aversive event (the unconditioned stimulus; US) is learned (CS-US association). During a subsequent extinction-learning phase, the CS+ is repeatedly presented without the presentation of the US, which is thought to result in the formation of a new association between the CS and omission of the US (CS-no US association) [8]. This process is similar to exposure therapy, during which feared stimuli are presented without the occurrence of anticipated negative outcomes [6]. Modern accounts strongly suggest that extinction learning represents a form of inhibitory learning [9], as the new association (CS-no US) suppresses the original threat association that remains intact [10]. The learning process is thought to be driven by expectancy violations, also known as prediction errors, that occur during the omission of the expected US. These expectancy violations are thought to be high during initial US

¹Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, EURON, Maastricht University Medical Centre, Maastricht, the Netherlands; ²Faculty of Psychology and Educational Sciences, Laboratory of Biological Psychology, University of Leuven, Leuven, Belgium; ³Institute for Mental Health Care Eindhoven (GGZE), Eindhoven, the Netherlands; ⁴Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; ⁵Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands; ⁶Department of Psychosis Studies, Institute of Psychiatry, King's College London, King's Health Partners, London, UK and ⁷Faculty of Psychology, Center for Experimental and Learning Psychology, University of Leuven, Leuven, Belgium

Correspondence: Iris Lange (i.lange@maastrichtuniversity.nl)

Received: 8 February 2019 Revised: 13 July 2019 Accepted: 19 July 2019

Published online: 28 July 2019

omissions at the beginning of extinction learning, as the US is still expected. The expectancy violations however decrease over extinction learning trials, as the safe outcome (no US) is increasingly anticipated and thereby the difference between the expected and actual outcome decreases [11]. Furthermore, expectancy violations during the therapeutic process are thought to be crucial for a successful therapy response [12, 13].

Insight into the neurobiological mechanisms underlying exposure therapy largely stems from preclinical studies and human neuroimaging studies on extinction learning and recall. Studies capturing time-related changes during extinction learning have established a critical role of the infralimbic cortex or human homologous ventromedial prefrontal cortex (vmPFC) in extinction, with the vmPFC becoming increasingly activated over the extinction learning phase [14–17]. In addition, the connection with the hippocampus has been implicated in the consolidation and retrieval of the extinction memory (for a review: [18]). Increased engagement of the vmPFC during extinction recall has been associated with reduced psychophysiological responses to the extinguished CS+ [17, 19]. Individuals with anxiety disorders and post-traumatic stress disorder (PTSD) however may suffer from deficiencies in vmPFC-mediated inhibition, resulting in difficulties with extinction learning and recall [20–22]. Furthermore, regions involved in prediction error signaling are thought to play a role in the inhibitory extinction learning process [11]. Although receiving less attention in models of fear extinction learning, the nucleus accumbens (NAcc) has been pointed out as one of the key regions involved in prediction error signaling. Via phasic dopamine release from the ventral tegmental area (VTA), the NAcc becomes activated during positive prediction errors, which occur during unexpected obtaining of rewards as well as omissions of expected US [23–26]. Next to the NAcc, it has been demonstrated that the vmPFC becomes activated in response to US omission during extinction learning [25]. Whether individuals with anxiety disorders and PTSD also show deficiencies in extinction learning-related prediction error signaling is however currently unknown.

The evaluation of predictive validity of extinction learning as a model for exposure therapy, i.e., whether individual differences in extinction learning and recall predict exposure therapy outcome, has only been explored by a few studies [27, 28]. In social anxiety disorder, it was shown that both recruitment of the vmPFC during the early phase of extinction learning, and lower differentiation in valence between the CS+ and conditioned safety stimulus (CS-) predicted better therapy response [27]. Furthermore, a study in spider fearful participants found that reduction of US expectancy and some physiological indices of stronger extinction learning predicted response to a treatment-analogue of exposure therapy [28]. This study additionally explored the link between extinction recall and therapy outcome. When retention of the extinction memory is reduced, or generalization of the extinction memory is limited, a non-optimal therapy outcome or relapse may be expected [4]. Unexpectedly, there were some indications that better retention of the extinction memory was related to a lower response to this treatment-analogue of exposure therapy [28], a finding that would need further examination.

Although extinction learning has been regarded as a core mechanism underlying exposure therapy, it has not been examined how specific the predictive value of extinction learning is for exposure therapy success. More specifically, extinction learning could also be a common indicator of symptom improvement and therefore unspecific to any therapeutic method. Other therapeutic strategies for anxiety, such as relaxation strategies or cognitive therapy are however not particularly assumed to rely on extinction learning [29]. Investigating the specificity of extinction learning as underlying mechanism of exposure therapy requires the inclusion of an additional therapy

arm to examine whether individual differences in extinction learning predicts response to this second intervention.

The research questions of current study are the following: (1) Does inter-individual variability in self-reported US expectancy and neural activation in the vmPFC and NAcc during extinction learning and extinction recall predict exposure therapy outcome? (2) How specific are these predictions for response to exposure therapy, compared to response to an alternative intervention? The current study included individuals with a phenotypically clear disorder (i.e., specific phobia for spiders). Participants were randomized to either group-based exposure therapy or progressive muscle relaxation (PMR). Although exposure therapy is the gold standard for specific phobia and outperforms other therapies in terms of symptom reductions, PMR can be regarded as an active control condition which can significantly reduce phobia symptoms [30–32]. Based on prior evidence [28, 33], we hypothesized that a better response to exposure therapy, but not PMR, is predicted by higher vmPFC activation during early extinction learning and extinction recall, by increased vmPFC and NAcc activation during early US omissions, and by greater reductions in US expectancy during extinction learning.

MATERIALS AND METHODS

Participants

Participants aged 16–25 years were recruited via local advertisements and included as part of a larger randomized-controlled trial in the south of Limburg, The Netherlands (Smartscan, Dutch Trial Register Number: NTR380, <https://www.trialregister.nl/trial/3662>; dates 10/2013–05/2018), (see Fig. 1 for flow chart). The sample size was determined based on a power calculation aiming to find a significant difference between exposure therapy and PMR in symptom reductions. Participants were required to meet the DSM-IV criteria for a specific phobia as confirmed by the MINI International Neuropsychiatry Interview [34]. Exclusion criteria were any other current psychiatric diagnosis or treatment, and any past diagnosis other than an anxiety disorder, a history of neurological disease, use of psychotropic medication, and MRI contra-indications. The study was approved by the local medical ethics committee. All participants received full information on study procedures and provided informed consent before study onset. Parental consent was additionally required for minors (<age 18 years). Forty-five participants completed data for the current study.

Study design and randomization

The design was a double-blind RCT. Participants were unaware of the active condition. During baseline (T0), participants were asked to complete the symptom questionnaires (see 'Psychopathology measures') and then underwent the 3-day fear learning, extinction learning, and extinction recall paradigm (see 'fMRI extinction learning and recall task'). Whenever enough participants completed the baseline to form a group for the group sessions (3–5 individuals), participants were as a group randomly assigned to, and received either group-based exposure therapy ($n = 25$) or PMR ($n = 20$), 4 weeks after T0. Allocation was done via a random generated list, and performed by an independent researcher to control for blinding of assessors. Allocation concealment was ensured by having the independent researcher only informing the therapists (via email). The Fear of Spiders Questionnaire (FSQ) was administered again 4 weeks post-therapy (T1).

Interventions

Exposure therapy consisted of a group-based single session (3–5 participants) under the guidance of a psychiatrist (KL). In a hierarchical manner, participants were exposed to spider-relevant stimuli and actual spiders through 20 steps, starting with drawing a spider and ending with holding a large house spider on the

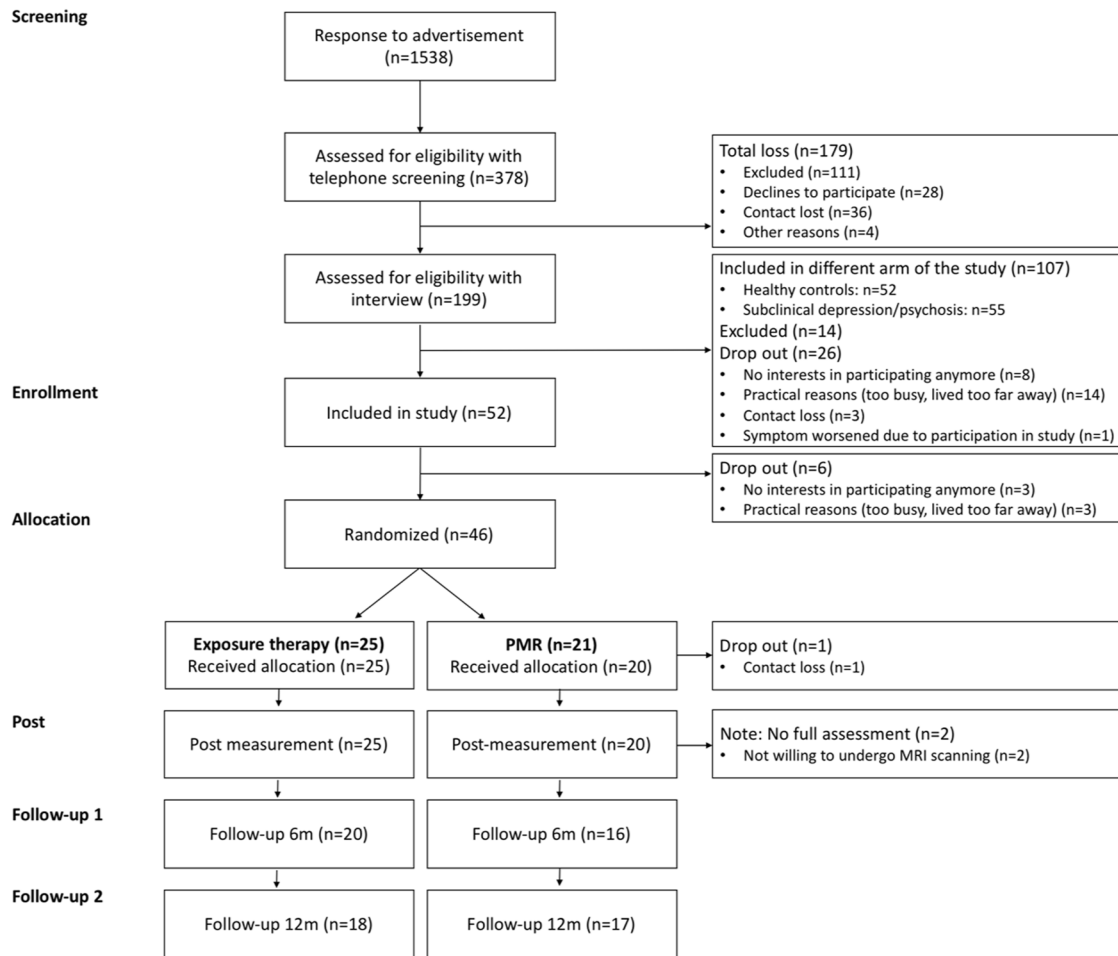


Fig. 1 Flow chart of the randomized controlled trial data used for the current study

hand. Similarly, PMR consisted of a group-based (3–5 participants) single session led by a psychologist (IL). The session started with deep breathing exercises, followed by instructions regarding tightening and relaxation of several muscle groups. Further instructions were given using standard audiotape, and the session ended when the entire relaxation protocol was finished. No references were made to spiders or spider-relevant stimuli. At the end of either intervention, participants were encouraged to continue doing either exposure or relaxation exercises at home, depending on the group they were in.

Psychopathology measures

The main outcome of the RCT was spider phobia severity, assessed with the FSQ [35]. The FSQ is a self-report questionnaire, has 18 items, with varying scores between 18 and 126. Additionally, to assess whether intervention groups did not differ on depression scores and trait anxiety, the Inventory for Depressive Symptomatology – Self-report (IDS-SR) [36] and Spielberger’s state and trait anxiety scales [37] were assessed at baseline (T0).

fMRI extinction learning and recall task

A validated fMRI task was used for the current study. Results obtained from the current sample and healthy controls have been previously published [15]. The task was presented via the Presentation software package (Albany, CA, USA). Detailed task information can be found in the Supplemental Information.

In short, the task consisted of a three-day protocol: (day 1) pre-conditioning and fear conditioning, (day 2) extinction learning; (day 3) extinction recall (Fig. 2). During the pre-conditioning phase

on Day 1, geometrical shapes were shown (seven rings or rectangles of parametrically increasing sizes, and a triangle). The pre-conditioning phase was immediately followed by a fear conditioning phase. During the fear conditioning phase, the smallest circle/rectangle served as the conditioned threat stimulus (CS+) and the largest circle/rectangle as the conditioned safety stimulus (CS) for half of the participants. For the other half, the smallest and largest circle/rectangle were respectively the CS– and CS+. In addition, the triangle served as a second conditioned safety stimulus (vCS–). The CS+ co-terminated with the US (66% reinforcement schedule), a 200 ms electrical shock applied to the left ankle. The following day (Day 2), the CS+, CS– and vCS– were shown during extinction learning. On Day 3 all stimuli CS (CS+, CS–, vCS–) were shown during the extinction recall phase. In addition to the CS, the five intermediate rings/rectangles were shown, which served as generalization stimuli (GS). GS1 was most similar to the CS+, GS5 was least similar to the CS+. These stimuli were added to measure generalization of extinction. Stimuli were presented 12 times per task phase. During each task phase, participants reported US expectancy four times for each stimulus on a four-point scale (1 = no risk; 2 = low risk; 3 = moderate risk; 4 = high risk of receiving a shock). Before and after each task phase, participants were asked to rate the level of fear for each stimulus on a visual analogue scale (VAS (0 (not fearful) to 100 (very fearful))).

Analyses

Self-report data. Data were analyzed with SPSS version 24.0. Intervention effects were explored via an Analysis of Covariance

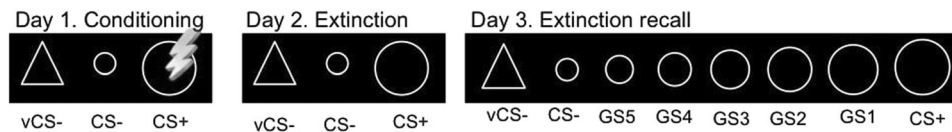


Fig. 2 Neuroimaging fear and extinction learning task: Stimuli per experimental phase. One task version (large circle is CS+) out of four possible options is presented. CS+ = conditioned threat stimulus; (v)CS- = conditioned safety stimulus; GS = generalization stimulus

with post-therapy FSQ scores as outcome, group as between-subject factor, and baseline FSQ as a covariate. Analyses on the main effects of the fear learning and extinction task across the entire sample have been previously reported [15] and can be found in the Supplemental Information. All analyses were corrected for age and sex.

To test our hypothesis, the relationship between extinction learning/recall outcomes and therapy outcome was examined with regression analyses. Therapy outcome was the FSQ score at post-therapy (T1), with all analyses controlling for the baseline FSQ score (T0). Extinction learning in self-report data was defined as following: (1) change in response to CS+ from pre-to-post extinction (pre minus post), 2. CS+ >vCS- differentiation at the end of extinction. Extinction recall was defined as following: (1) CS+ >vCS- differentiation, and (2) CS+ >average(GS1-5) differentiation at the beginning of extinction. These outcomes were calculated both for fear ratings and US expectancy scores. Significant analyses were followed-up by interaction analyses to examine the specificity of these results for the exposure therapy group, by evaluating the interaction between group (exposure, relaxation) and extinction learning/recall outcomes on post-therapy FSQ scores. Analyses were controlled for multiple comparisons by Bonferroni correction, separately for extinction learning and extinction recall. The *p*-value was adjusted for the number of self-report outcomes for both measures (fear and US expectancy), and therefore a *p* < 0.0125 was considered statistically significant.

MRI analyses. Information regarding MRI acquisition, MRI data preprocessing, first-level analyses, and analyses regarding main task effects can be found in the Supplemental Information. To investigate the link between neural activations during early extinction learning and exposure therapy outcome, the following contrasts of interest were defined: (1) early CS+ >early vCS-; (2) early US omission > early vCS- offset. Early extinction was defined as the first four trials of the extinction learning phase. Mean percent signal change (PCS) from these extinction contrasts images was extracted in functional regions of interests: (1) vmPFC (all contrasts), (2) NAcc (only offset contrast) (see Supplemental Information). Contrasts of interest of the extinction recall were the following: (1) early CS+ >early vCS-, and (2) early CS+ >early GS (1-5). The contrasts were based on the first four trials of the extinction recall phase. PCS for the recall contrasts were extracted within an anatomical mask of the vmPFC, as defined by the probabilistic FSL's Harvard-Oxford atlases thresholded at 50%. Subsequently, regression analyses were conducted to test the association between PCS within the ROIs and post-therapy FSQ scores, controlling for baseline FSQ, age, and sex. Significant analyses were followed-up by interaction analyses to examine whether the results were specific for the exposure therapy group. More specifically, these analyses the interaction between group (exposure, PMR) and extinction learning/recall-related activations on post-therapy FSQ scores. Exploratory analyses on late extinction contrasts (1) late CS+ >late vCS-; (2) late US omission >late vCS- offset) are reported in the Supplemental Information. These analyses were also controlled for multiple comparisons, taking into account the number of contrasts and the number of regions, separately for extinction learning and extinction recall. Therefore respectively for extinction learning

and extinction recall, *p* < 0.017 and *p* < 0.025 was considered statistically significant.

Additional exploratory whole-brain regression analyses were carried out in order to examine the association between extinction learning/recall activations and intervention outcome, corrected for multiple comparisons at the cluster level using Gaussian random field theory (GRF); (*Z* > 2.3, cluster-wise *p* < 0.05) using FSL's FMRIB's Local Analysis of Mixed Effects (FLAME1), of which the results are reported in the Supplemental Information.

RESULTS

Participants

Baseline demographic and clinical measures are shown in Table 1. No differences between exposure therapy and PMR were found, neither on the demographic measures (age: *t*(1, 43) = 0.48, *p* = 0.63; sex: $\chi^2(1) = 1.66$, *p* = 0.20; educational level: $\chi^2(1) = 1.36$, *p* = 0.24), nor on baseline psychopathology outcomes (FSQ: *t*(1, 43) = -1.20, *p* = 0.24; trait anxiety *t*(1, 43) = -0.76, *p* = 0.45; depression: *t*(1, 43) = -0.86, *p* = 0.40), nor on US intensity used during the conditioning experiment (*t*(1, 43) = 0.28, *p* = 0.78).

Manipulation checks

Therapy response. Post-therapy FSQ phobic symptoms were lower in the exposure therapy group than in the PMR group, as reflected by an ANCOVA controlling for baseline symptom severity (*F*(1, 43) = 8.13, *p* = 0.007; Cohen's *d* = 0.90; 95% confidence interval: 0.27-1.51). Post hoc paired-samples *t*-tests shown that both interventions significantly reduced phobic symptoms (PMR: *t*(1, 19) = 2.96, *p* = 0.008; exposure therapy: *t*(1, 24) = 8.45, *p* < 0.001) (Fig. 3).

Experimental task-self-report and neural outcomes. The experimental task was able to induce differential fear learning and extinction learning, with expected activations in neural circuitry, including the vmPFC and left NAcc during extinction learning (See Supplemental Information for further details).

Hypotheses testing

Extinction learning. Imaging: ROI analyses revealed that lower activation in the vmPFC during early extinction (early CS+ >early vCS-) predicted better exposure therapy outcome (lower post-therapy FSQ scores) ($\beta = 36.97$, *p* = 0.01) (Fig. 4a). Follow-up analyses revealed that this effect was mainly driven by increased response to the vCS- being associated with lower post-therapy FSQ scores ($\beta = -20.32$, *p* = 0.05), and not by response to the CS+ ($\beta = 2.73$, *p* = 0.81). In addition, activations during early US omissions were associated with exposure therapy outcome. More specifically, better exposure therapy outcome was related to increased vmPFC activation ($\beta = -278.06$, *p* = 0.001) and marginally to increased left NAcc activation ($\beta = -34.36$, *p* = 0.06) during early US omission (early US omission > early vCS- offset) (Fig. 4b, c). The effect in the vmPFC was driven by increased US omission-related activation being related to lower post-therapy FSQ scores ($\beta = -156.24$, *p* = .01) but not response to vCS- offset ($\beta = 46.65$; *p* = 0.63). Further interaction analyses testing whether the strength of the association between post-therapy FSQ scores and extinction learning-related activations (in vmPFC and left NAcc) differed between exposure therapy and PMR revealed that

Table 1. Overview of demographic and clinical variables for all participants, and for the exposure therapy group and progressive muscle relaxation (PMR) group separately

	Overall n = 45	Exposure n = 25	PMR n = 20
Age	20.6 (2.4)	20.4 (2.4)	20.8 (2.4)
Gender M/F	4/45	1/24	3/17
Educational level high/medium	40/5	21/4	19/1
FSQ	99.2 (15.4)	101.6 (14.5)	96.1 (16.3)
Trait Anxiety	30.4 (8.0)	31.2 (7.6)	29.4 (8.6)
Depression	8.5 (10.0)	9.6 (10.2)	7.1 (9.9)
US intensity	21.3 (15.5)	20.7 (15.0)	22.1 (16.7)

A high educational level corresponds to Bachelor's level or higher; a mid educational level to vocational education or higher
M male, F female, FSQ fear of Spiders Questionnaire, US unconditioned stimulus

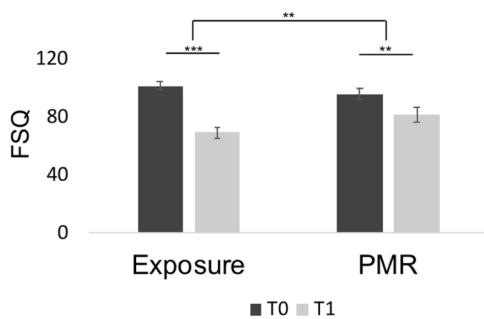


Fig. 3 Baseline (T0) and post-therapy (T1) phobia symptom scores for each intervention measured with the Fear of Spiders Questionnaire (FSQ). ** $p < 0.001$; *** $p < 0.001$. Error bars represent standard error of the mean

this was the case for US omission-related activation in the vmPFC ($\beta = -371.62, p = 0.02$), but not for vmPFC activation during early extinction ($\beta = 33.88, p = 0.14$), nor for US omission-related activation in the left NAcc ($\beta = -30.78, p = 0.18$). Follow-up analyses in the PMR group only, revealed no significant associations between post-therapy FSQ scores and activation in any of the ROIs (vmPFC during early CS+ > early vCS-: $\beta = 4.16, p = 0.83$; vmPFC during early US omission > early vCS- offset: $\beta = 87.79, p = 0.55$; left NAcc during early US omission > early vCS- offset: $\beta = -5.96, p = 0.70$).

Whole-brain analyses. Self-report: Exposure therapy outcome was not related to any self-report indices of extinction learning, as post-therapy FSQ scores were not associated with reductions in task fear scores ($\beta = 1.27, p = 0.74$) or US expectancy ($\beta = -0.40, p = 0.83$) over the extinction learning phase, or with differential CS+ > vCS- fear ($\beta = 3.63, p = 0.48$) or US expectancy scores ($\beta = -0.13, p = 0.35$) at the end of extinction learning.

Extinction recall. Imaging: Activation in the vmPFC during extinction recall was not associated with exposure therapy outcome (CS+ > vCS-: $\beta = 7.39, p = 0.60$; CS+ > average(GS1-5): $\beta = -2.13, p = 0.86$).

Self-report: Exposure therapy outcome was not related to any self-report extinction recall indices, as no associations were observed between post-therapy FSQ scores and differential CS+ > vCS- fear ($\beta = -0.16, p = 0.23$) and US expectancy scores ($\beta = -2.80, p = 0.44$), nor between post-therapy FSQ scores and

CS+ > average(GS1-5) differentiation in fear ratings ($\beta = -0.65, p = 0.88$) and US expectancy ($\beta = -3.19; p = 0.50$).

DISCUSSION

Based on the assumption that extinction learning represents the core principle underlying exposure therapy, the current study investigated whether inter-individual differences in extinction learning and extinction recall were predictive of response to exposure therapy in specific phobia relative to an alternative intervention (i.e., PMR). Exposure therapy was more effective in reducing phobia symptoms compared to PMR. As expected, increased vmPFC activation during early US omissions specifically predicted better response to exposure therapy and not to PMR. Exposure therapy response was also associated with increased recruitment of the vmPFC during safety stimulus (vCS-) processing but not to threat stimulus (CS+) processing as initially hypothesized. In addition, there were some trending positive associations between left NAcc activation during early US omissions and therapy outcome. These latter two findings in the exposure therapy group were not significantly different from the PMR group. Contrary to our hypotheses, neural responses during extinction recall and self-report measures (i.e., fear and US expectancy) were not predictive of exposure therapy outcome.

The current study explored neural prediction error signaling by modeling US omissions during extinction learning, as prediction errors are thought to drive the inhibitory extinction learning process. This observation has received much attention in the field of experimental and clinical psychology, but neurobiological research on US omission-related prediction error signaling during extinction learning is largely lacking. Our results showed that higher vmPFC during early US omissions were specifically related to greater symptom reductions following exposure therapy but not PMR. Both animal and human research have demonstrated that the vmPFC is critically involved in extinction learning and retrieval of the extinction memory [18]. A role of the vmPFC in prediction error signaling during early extinction learning has also been previously reported [25], next to its general role in safety signal processing and fear inhibition [38]. Our data showed, for the first time, a link between neural prediction error signaling during extinction learning and response to exposure therapy and provide evidence for the specificity of this link. The results suggest that individuals with increased neural prediction error-related responses during threat omission are those individuals that respond best to exposure therapy. Behavioral or neurobiological strategies (e.g., repeated transcranial magnetic stimulation) that respectively enhance safety processing or enhance vmPFC functioning during threat omissions may be a target for improving response to exposure therapy.

Our neural results further showed that higher vmPFC activation during safety stimulus processing (CS-), at the beginning of extinction learning, was associated with exposure therapy outcome. This result was unexpected, as it is at odds with a previous study [27], showing that greater vmPFC activation during threat stimulus processing at early extinction (CS+ > CS- contrast) was associated with exposure therapy response. Our results therefore suggest that adaptive safety stimulus processing, instead of the threat stimulus processing, may represent an essential mechanism in therapy response. Furthermore, although no significant associations were found between this vmPFC activation and symptom reductions within the PMR group, the intervention groups (exposure, PMR) did not statistically differ in the strength of the associations between these neural activations and symptom reductions. Therefore, it is unclear to what degree this result is specific to the exposure therapy group. During exposure therapy, phobic individuals are often exposed to several feared stimuli that are similar to the CS+ but actually convey safety [39]. Appropriate inhibition of conditioned fear responses to these

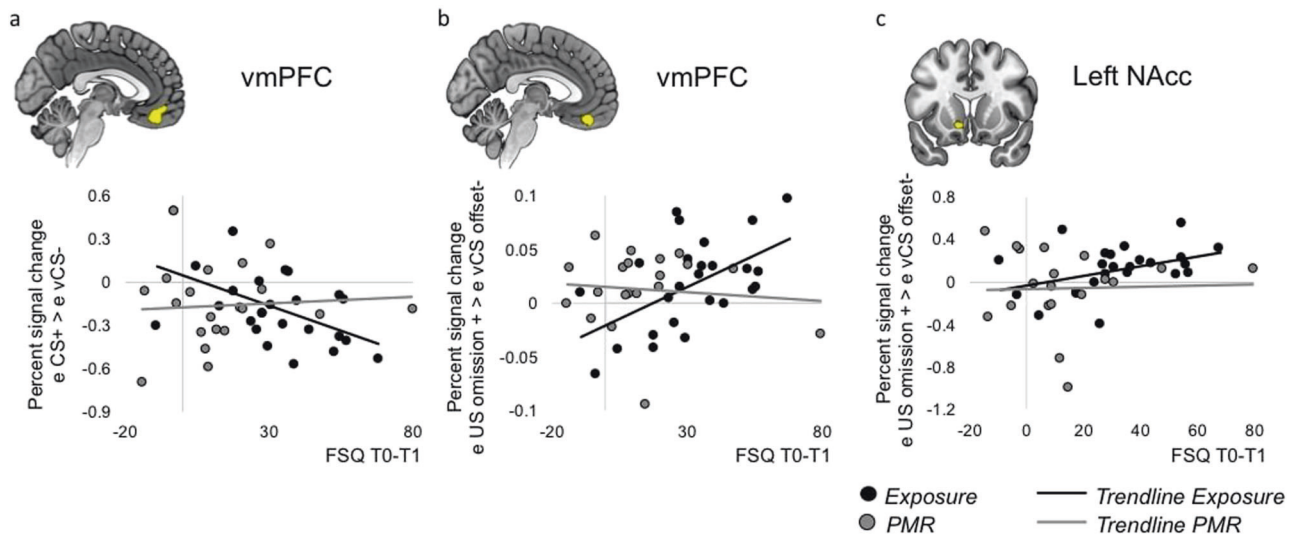


Fig. 4 Associations between extinction-related activation and symptom reductions (FSQ T0-T1) following exposure therapy and PMR. **a** Association between symptom reductions and vmPFC activation during early extinction (CS+ > vCS-), **b** Association between symptom reductions and vmPFC activation during early omissions of the unconditioned stimulus (US) (US omission > vCS- offset), **c** Association between symptom reductions and left nucleus accumbens (NAcc) activation during early US omissions (US omission > vCS- offset). T0 = baseline; T1 = post-therapy; e = early

stimuli is important in adaptive anxiety [40], and extinction of fear responses to generalization stimuli is considered to be critical for exposure therapy success [41]. During early extinction learning, the meaning of the conditioned safety stimulus could be considered ambiguous and potentially threatening. Our results could therefore indicate that exposure success is related to the degree to which fear responses to a safety stimulus are inhibited or extinguished. Furthermore, our results underline the importance of including an additional condition in mechanistic therapy research, to be able to establish the specificity of findings for a particular therapeutic technique of interest.

There were also some small indications at the trend level showing that increased left NAcc activation during early (and late) US omission was related to therapy outcome but this effect also not specific to exposure therapy. The NAcc is considered as one of the core regions involved in reward prediction error signaling. Previous research has shown that unexpected omission of an aversive stimulus, as occurs during extinction learning, also activates this region [24]. Our study adds to the emerging literature, mostly in rodents so far, on the role of the NAcc in US omission signaling during extinction learning [42, 43]. As our therapy prediction results in the NAcc were only at trend level, these merely cautiously suggest that individuals with enhanced prediction error signaling during extinction learning may respond better to exposure therapy. Such an interpretation would be in line with behavioral and experimental research suggesting that larger US expectancy violations result in stronger inhibitory learning and therefore in an increased efficacy of exposure therapy [6].

Contrary to what we hypothesized, exposure therapy outcome was not related to any indices of extinction recall. Exposure therapy response has been suggested to rely on both extinction learning and extinction recall mechanisms [18], two processes that might be rather independent. Although this is the first study linking neural extinction recall measures to exposure therapy outcome, a previous study measuring self-report and physiology during extinction recall also found no clear indication that these indices predicted therapy outcome [28]. In addition, the current study also did not find self-reported extinction learning data to be predictive of symptom reductions. Previous studies report mixed findings on predictive effects of

self-report extinction indices on exposure therapy outcome. For example, one study found that both valence and arousal ratings during extinction learning were not related to outcome to cognitive-behavioral therapy in anxious children [44]. On the contrary, a recent study showed that a lower differentiation in valence scores between CS+ and CS- at the end of extinction was associated with greater reductions in social anxiety symptoms after exposure therapy [27]. Similarly, a second study has reported that reduced CS+ > CS- differentiation in US expectancy scores during early extinction learning was associated with greater symptom reductions after an exposure therapy analogue for spider phobia [28]. Yet, in these two latter studies, no consistent associations were found between all extinction-related self-report indices and therapy response outcome measures [27, 28]. This may indicate that a dissociation exists between different measures of threat processing, as previously reported [45]. Lack of stable replication of findings may be further explained by variation in protocols and calculation of extinction learning and therapy outcome indices.

Some limitations should be pointed out. The sample mainly consisted of female participants, congruent with the greater prevalence of specific phobia in women compared to men [46]. Our results may therefore not generalize readily to males with a phobia. Secondly, the phobia symptom severity scores were obtained 4 weeks before and after therapy. It could have been informative to examine phobia symptom change immediately after exposure therapy and to track within-session changes in fear and threat expectancies. Thirdly, no physiological measures were obtained during the experimental task, which could have provided an additional measure of threat responding. Fourthly, our experimental task did not include phobia-relevant stimuli. Using such stimuli might have increased the ecological validity of our experimental task. Fifthly, although specific phobia represents a model disorder for studying fear-related psychopathology in terms of its high prevalence, frequently lacking comorbidities and pharmacotherapy use, and high symptom homogeneity, the generalizability of the results to other anxiety disorders needs to be further explored. Lastly, the number of CS+ presentations (12 in total) in the extinction learning phase was relatively small which may have limited occurrence of inhibitory learning to occur. The

extinction-learning phase did however result in an overall reduction in fear and US expectancy responses, recruitment of vmPFC, and in sufficient individual variation. Future replication research is needed to confirm our findings. Furthermore, exposure therapy, as applied in daily clinical practice likely does not only consist of exposure/extinction, but also includes other processes such as active approach behaviors, counterconditioning and fear management skills, which may also contribute to the therapeutic effect [7, 47]. Future research could focus on these mechanisms in relation to therapeutic outcome and their interaction with extinction learning.

In conclusion, results of the current study suggest that individuals with stronger prediction error-related signaling in the vmPFC during extinction learning specifically show a better response to exposure therapy. Furthermore, how strongly the vmPFC processes safety signals during extinction learning could be predictive of exposure therapy response, but the specificity of this association is less clear. Results suggest that boosting the extinction circuitry may be a target for improving exposure therapy success. The present results further question the validity and usefulness of self-reported laboratory measures (i.e., fear and US expectancy) during extinction learning and recall as experimental proxies of exposure therapy response.

FUNDING AND DISCLOSURE

This study was funded by a research grant from Stichting De Weijerhorst and by a fellowship from the Hersenstichting granted to M. Wichers (Hersenstichting Nederland: 2012(1)-03). The authors report no biomedical financial interests or potential conflicts of interest.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41386-019-0467-8>).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Ougrin D. Efficacy of exposure versus cognitive therapy in anxiety disorders: systematic review and meta-analysis. *BMC Psychiatry*. 2011;11:200.
- Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cogn Ther Res*. 2012;36:427–40.
- Deacon BJ, Abramowitz JS. Cognitive and behavioral treatments for anxiety disorders: a review of meta-analytic findings. *J Clin Psychol*. 2004;60:429–41.
- Vervliet B, Craske MG, Hermans D. Fear extinction and relapse: state of the art. *Annu Rev Clin Psychol*. 2013;9:215–48.
- Holmes EA, Craske MG, Graybiel AM. A call for mental-health science. *Nature*. 2014;511:287.
- Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. *Behav Res Ther*. 2014;58:10–23.
- Scheveneels S, Boddez Y, Vervliet B, Hermans D. The validity of laboratory-based treatment research: bridging the gap between fear extinction and exposure treatment. *Behav Res Ther*. 2016;86:87–94.
- Bouton ME. Context and behavioral processes in extinction. *Learn Mem*. 2004;11:485–94.
- Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. *Behav Res Ther*. 2008;46:5–27.
- Bouton ME. Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol Bull*. 1993;114:80.
- Dunsmoor JE, Niv Y, Daw N, Phelps EA. Rethinking extinction. *Neuron*. 2015;88:47–63.
- Guzick AG, Reid AM, Balkhi AM, Geffken GR, McNamara JP. That was easy! expectancy violations during exposure and response prevention for childhood obsessive-compulsive disorder. *Behav Modification*. 2018;0145445518813624.

- Salkovskis PM, Hackmann A, Wells A, Gelder MG, Clark DM. Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: a pilot study. *Behav Res Ther*. 2007;45:877–85.
- Sotres-Bayon F, Quirk GJ. Prefrontal control of fear: more than just extinction. *Curr Opin Neurobiol*. 2010;20:231–35.
- Lange I, Goossens L, Bakker J, Michielse S, Marcelis M, Wichers M, et al. Functional neuroimaging of associative learning and generalization in specific phobia. *Prog Neuro-Psychopharmacol Biol Psychiatry*. 2018;89:275–85.
- Schiller D, Kanen JW, LeDoux JE, Monfils M-H, Phelps EA. Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proc Natl Acad Sci*. 2013;110:20040–5.
- Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry*. 2007;62:446–54.
- Milad MR, Quirk GJ. Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol*. 2012;63:129–51.
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron*. 2004;43:897–905.
- Milad MR, Orr SP, Lasko NB, Chang Y, Rauch SL, Pitman RK. Presence and acquired origin of reduced recall for fear extinction in PTSD: results of a twin study. *J Psychiatr Res*. 2008;42:515–20.
- Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, et al. Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol Psychiatry*. 2009;66:1075–82.
- Dibbets P, van den Broek A, Evers EA. Fear conditioning and extinction in anxiety- and depression-prone persons. *Memory*. 2015;23:350–64.
- Hart AS, Rutledge RB, Glimcher PW, Phillips PE. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci*. 2014;34:698–704.
- Baliki MN, Geha PY, Fields HL, Apkarian AV. Predicting value of pain and analgesia: nucleus accumbens response to noxious stimuli changes in the presence of chronic pain. *Neuron*. 2010;66:149–60.
- Gerlicher A, Tüscher O, Kalisch R. Dopamine-dependent prefrontal reactivations explain long-term benefit of fear extinction. *Nat Commun*. 2018;9:4294.
- Day JJ, Carelli RM. The nucleus accumbens and Pavlovian reward learning. *Neuroscientist*. 2007;13:148–59.
- Ball TM, Knapp SE, Paulus MP, Stein MB. Brain activation during fear extinction predicts exposure success. *Depress Anxiety*. 2016;34:257–66.
- Forcadell E, Torrents-Rodas D, Vervliet B, Leiva D, Tortella-Feliu M, Fullana MA. Does fear extinction in the laboratory predict outcomes of exposure therapy? A treatment analog study. *Int J Psychophysiol*. 2017;121:63–71.
- Arch JJ, Craske MG. Acceptance and commitment therapy and cognitive behavioral therapy for anxiety disorders: different treatments, similar mechanisms? *Clin Psychol: Sci Pract*. 2008;15:263–79.
- Öst L-G, Johansson J, Jerremalm A. Individual response patterns and the effects of different behavioral methods in the treatment of claustrophobia. *Behav Res Ther*. 1982;20:445–60.
- Wolitzky-Taylor KB, Horowitz JD, Powers MB, Telch MJ. Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clin Psychol Rev*. 2008;28:1021–37.
- Gilroy LJ, Kirkby KC, Daniels BA, Menzies RG, Montgomery IM. Controlled comparison of computer-aided vicarious exposure versus live exposure in the treatment of spider phobia. *Behav Ther*. 2000;31:733–44.
- Ball TM, Knapp SE, Paulus MP, Stein MB. Brain activation during fear extinction predicts exposure success. *Depress anxiety*. 2017;34:257–66.
- Overbeek I, Schruers K, Griez E. Mini international neuropsychiatric interview: Nederlandse versie 5.0. O. DSM-IV [Dutch version]. Maastricht, The Netherlands: Universiteit Maastricht; 1999.
- Szymanski J, O'Donohue W. Fear of spiders questionnaire. *J Behav Ther Exp Psychiatry*. 1995;26:31–34.
- Rush AJ, Carmody T, Reimtz PE. The Inventory of Depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *Int J Methods Psychiatr Res*. 2000;9:45–59.
- Spielberger CD State-Trait anxiety inventory. Wiley Online Library; 2010.
- Harrison BJ, Fullana MA, Via E, Soriano-Mas C, Vervliet B, Martínez-Zalacain I, et al. Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *NeuroImage*. 2017;152:12–8.
- Struyf D, Hermans D, Vervliet B. Maximizing the generalization of fear extinction: exposures to a peak generalization stimulus. *Behav Res Ther*. 2018;111:1–8.
- Lissek S, TOWARD AN. Account of clinical anxiety predicated on basic, neurally mapped mechanisms of pavlovian fear-learning: the case for conditioned over-generalization. *Depress Anxiety*. 2012;29:257–63.
- Craske MG, Hermans D, Vervliet B. State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philos Trans R Soc B*. 2018;373:20170025.

42. Luo R, Uematsu A, Weitemier A, Aquili L, Koivumaa J, McHugh TJ, et al. A dopaminergic switch for fear to safety transitions. *Nat Commun.* 2018;9:2483.
43. Holtzman-Assif O, Laurent V, Westbrook RF. Blockade of dopamine activity in the nucleus accumbens impairs learning extinction of conditioned fear. *Learn Mem.* 2010;17:71–5.
44. Waters AM, Pine DS. Evaluating differences in Pavlovian fear acquisition and extinction as predictors of outcome from cognitive behavioural therapy for anxious children. *J Child Psychol Psychiatry.* 2016;57:869–76.
45. Sevenster D, Beckers T, Kindt M. Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiol Learn Mem.* 2012;97:338–45.
46. Fredrikson M, Annas P, Fischer H, Wik G. Gender and age differences in the prevalence of specific fears and phobias. *Behav Res Ther.* 1996;34:33–9.
47. Hofmann SG. Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. *Clin Psychol Rev.* 2008;28:199–210.