PCN Psychiatry and Clinical Neurosciences

REGULAR ARTICLE

A double-blind trial of decoded neurofeedback intervention for specific phobias

Cody A. Cushing, PhD [1*] Hakwan Lau, PhD,[2] Mitsuo Kawato, PhD,[3,4] Michelle G. Craske, PhD[1] and Vincent Taschereau-Dumouchel, PhD[5,6*]

Aim: A new closed-loop functional magnetic resonance imaging method called multivoxel neuroreinforcement has the potential to alleviate the subjective aversiveness of exposure-based interventions by directly inducing phobic representations in the brain, outside of conscious awareness. The current study seeks to test this method as an intervention for specific phobia.

Methods: In a randomized, double-blind, controlled single-university trial, individuals diagnosed with at least two (one target, one control) animal subtype-specific phobias were randomly assigned (1:1:1) to receive one, three, or five sessions of multivoxel neuroreinforcement in which they were rewarded for implicit activation of a target animal representation. Amygdala response to phobic stimuli was assessed by study staff blind to target and control animal assignments. Pretreatment to posttreatment differences were analyzed with a two-way repeated-measures ANOVA.

Results: A total of 23 participants (69.6% female) were randomized to receive one ($n = 8$), three ($n = 7$), or five ($n = 7$) sessions of multivoxel neuroreinforcement. Eighteen ($n = 6$ each group) participants were analyzed for our primary outcome. After neuroreinforcement, we observed an interaction indicating a significant decrease in amygdala response for the target phobia but not the control phobia. No adverse events or dropouts were reported as a result of the intervention.

Conclusion: Results suggest that multivoxel neuroreinforcement can specifically reduce threat signatures in specific phobia. Consequently, this intervention may complement conventional psychotherapy approaches with a nondistressing experience for patients seeking treatment. This trial sets the stage for a larger randomized clinical trial to replicate these results and examine the effects on real-life exposure.

Clinical Trial Registration: The now-closed trial was prospectively registered at ClinicalTrials.gov with ID NCT03655262.

Keywords: *decoding, fMRI, neurofeedback, phobia, reinforcement.*

http://onlinelibrary.wiley.com/doi/10.1111/pcn.13726/full

Introduction

Fear-based disorders such as specific phobia are among the most difficult mental disorders to treat. The most widely empirically supported treatment is "exposure therapy," which involves direct exposure to fear-causing or panic-inducing stimuli.[1] This treatment is highly effective in reducing fear. However, conscious exposure to feared stimuli is a disturbing and unpleasant experience for the patient, leading to high rates of attrition.[2,3] Exposure treatment dropout rates can be as high as 70%, with 60% being unwilling to even start treatment.[1,3–5] As a result, only a small percentage of patients can actually benefit from an otherwise effective treatment.

Consequently, neurofeedback has been explored as a way of directly regulating brain activity in a number of mental health disorders.[6–12] A promising new functional magnetic resonance imaging (fMRI) method called multivoxel neuroreinforcement[13–15] has demonstrated the ability to lessen physiological defensive responses to both laboratory-conditioned fears and preexisting fears through a kind of "unconscious exposure."[16–19] Exposure treatments bypassing conscious awareness have shown promise in reducing fear responses in phobic individuals[20] owing to the dissociability of threat response and learning from subjective experience.[21,22] By using a machine-learning classifier (also referred to as a "decoder"), neuroreinforcement can be provided based on a specific stimulus category (e.g. spider) rather than average brain activity alone.[19] Critically, this results in no subjective discomfort for the patient, but yet can still lead to lasting reduction of fear.[7,8,18,23–25]

A decoder can be built for a patient with a phobia using brain data from a group of healthy controls for whom viewing repeated images of a target representation (e.g. spider) produces no fear reaction (Fig. 1). Training between-subject decoders this way enables "nonconscious exposure" in patients with phobias, without exposing them to feared stimuli. This surrogate data approach was explored in

[1] Department of Psychology, UCLA, Los Angeles, California, USA
[2] RIKEN Center for Brain Science, Wako, Japan
[3] Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, Kyoto, Japan
[4] XNef, Inc., Kyoto, Japan
[5] Department of Psychiatry and Addictology, Université de Montréal, Montreal, Québec, Canada
[6] Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montreal, Québec, Canada
* Correspondence: Email: ccushing@ucla.edu and vincent.taschereau-dumouchel@umontreal.ca

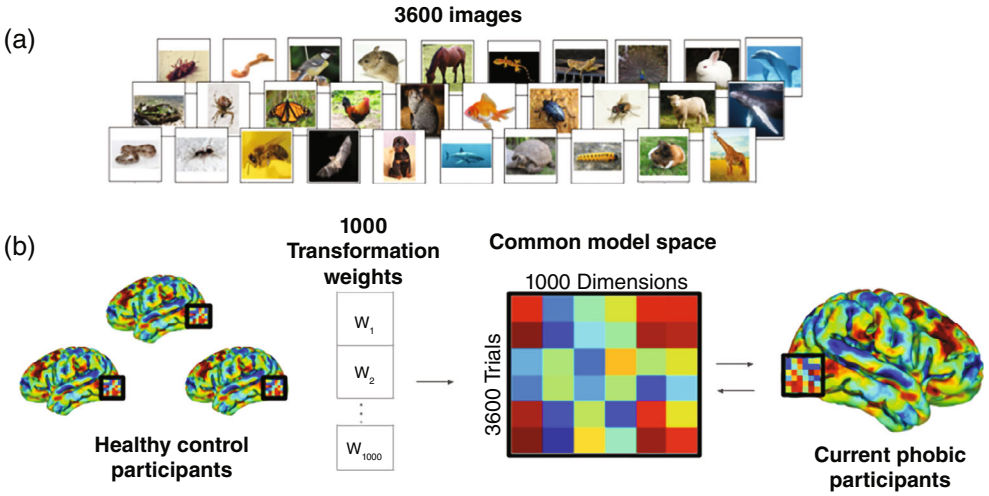

Fig. 1 Functional alignment of brain data into phobic participant brain using hyperalignment. (A) All participants complete a near-identical task in the functional magnetic resonance imaging scanner where 3600 images are rapidly viewed during 0.98-s presentations. Participants with phobia view happy human faces instead of their own phobic categories. Healthy controls view images from all categories. (B) Transformation parameters into the functionally aligned common model space are determined with phobic image trials withheld. Data from all participants for all categories (including phobic categories) are transformed into the common model space and then reverse transformed into the native space of the current phobic participant. A machine-learning classifier can then be trained on phobic images in the phobic participant's native brain space despite the participant never having personally viewed the images.

our previous proof-of-concept study,[17] but participants still saw the feared images during the decoder construction task. In the present study, we tested for the first time whether multivoxel neuroreinforcement can succeed using a decoder trained completely on surrogate data where the participant undergoing neuroreinforcement has never seen the feared images.

Here, we describe a preregistered (phase 1: https://osf.io/gjvmt?view_only=b6827aa394f143aeb29b99c095bd4183) double-blind, placebo-controlled clinical trial of this method as an intervention in a population with specific phobia. We preregistered five hypotheses (H1–H5). We hypothesized that amygdala responses (H1) and skin conductance responses (SCRs) (H2) to phobic stimuli would selectively decrease for the targeted phobia relative to the control phobia following neuroreinforcement. We focused on amygdala responding as our primary outcome due to its canonical role in learning and extinction of threat and fear responses.[26–30] In addition, we hypothesized that subjective fear ratings would stay the same following neuroreinforcement (H3), despite the predicted changes in physiological responses, based on our previous findings in a nonclinical population.[17] Secondarily, we introduced a modified affective Stroop task in which participants made rapid size judgments about phobic and neutral stimuli. In this task, we hypothesized that reaction times would be slower for phobic stimuli (H4i) and that following neuroreinforcement there would be a selective reduction in reaction times (H4ii) and amygdala responses (H4iii) in response to the targeted phobia category compared with the control phobia. Finally, we randomly assigned participants to receive either one, three, or five sessions of neuroreinforcement. We hypothesized that those receiving the most neuroreinforcement would demonstrate the largest effects (H5).

To anticipate, we did not manage to collect the full number of data ($N = 30$) as planned, due to pandemic-related circumstances. However, despite the reduced sample size ($N = 18$), our primary hypothesis about amygdala response reduction (H1) was confirmed. In addition, we found mixed evidence in support of secondary hypotheses concerning attentional capture by phobias in our novel affective Stroop task (H4). Unfortunately, due to circumstances outside our control, we lacked the statistical power to adequately assess the between-group differences for the amount of neuroreinforcement received (H5).

Methods

Trial design and participant screening

The current trial was designed as a randomized, within-subject, controlled, experimenter, and participant-blinded dose–response study with randomization to 1, 3, or 5 days of multivoxel neuroreinforcement and a primary end point of amygdala activation to targeted phobic stimuli compared with control phobic stimuli. The study protocol was approved by the institutional review board (IRB) at the University of California, Los Angeles, concordant with the provisions of the Declaration of Helsinki. Specific phobias were diagnosed using the *Anxiety Disorders Interview Schedule for DSM-IV*[31] in a diagnostic interview conducted by trained and reliability certified study staff. Details of diagnostic screening and control versus phobia grouping can be found in Supplemental Methods.

For multivoxel neuroreinforcement, 23 participants (mean age = 26.5 years [SD = 9.40 years], 69.6% female) with at least two specific animal phobias were enrolled. The informed consent of participants was obtained pursuant to the procedures of the IRB at the University of California, Los Angeles. Participants were randomly assigned to complete either one ($n = 8$), three ($n = 7$), or five ($n = 8$) days of multivoxel neuroreinforcement to determine the dose–response relationship with clinical outcomes. Of these 23 participants, two did not finish multivoxel neuroreinforcement (one because of technical issues, one because of scheduling issues). The principal investigator (M.G.C.) monitored the study on a day-to-day basis with prompt reporting of adverse events to the IRB, NIMH, and other agencies as appropriate. The following adverse events were monitored: deaths, suicide attempts, study dropout, psychiatric hospitalizations, and clinical deterioration as defined as emergent suicidal ideation or suicidal plan, development of serious substance abuse, or the emergence of a new psychiatric or medical diagnosis or behavior posing a significant risk to the subjects of others. Zero adverse events were recorded. Outcome analyses were performed on participants who completed the clinical trial per protocol. Of the 21 participants who completed multivoxel neuroreinforcement, one experienced nausea during tasks and was excluded from further analysis. Two participants did not complete the pre-post "fear test" task properly (closed their eyes or turned away in response to phobic images) for amygdala response (described below) and were excluded from analyses relevant to that task, leaving 18 participants ($n = 6$ each dosage group) for our primary analyses (H1, H2, H3, and H5). This cohort of 18 participants falls short of our original goal of 30 participants because our funding expired due to the shutdowns and recruitment difficulties resulting from the COVID-19 global pandemic. As a result, any further data collection was impossible. For secondary analysis of the affective Stroop task (H4), two of the 18 participants included in the fear test analysis did not complete the affective Stroop task, and one participant did not complete the fear test task properly, but did complete the affective Stroop task, resulting in 17 participants analyzed (H4).

Randomization and masking

On enrollment, participants were randomized to either one, three, or five sessions of neuroreinforcement using a random number generator by study coordinators with a 1:1:1 allocation. This randomization was not directly investigated during our primary analyses because of COVID-19 restrictions on data collection that limited power to detect

PCN Psychiatry and Clinical Neurosciences

between-group differences. However, the randomization group was controlled for as a covariate in all primary analyses. Neuroreinforcement itself was then controlled with a double-blind, within-subject placebo. Participants had at least two phobias with one phobia being used as the treatment target while another served as control. Assignment of the target and control phobias was performed automatically by a computer during data processing according to the procedures outlined in Supplemental Methods.

Experimenters were blinded during data collection by loading the target pattern automatically with computer software. Target and control pattern labels were automatically saved during decoder construction processing (outlined below) and stored as MATLAB variables to be automatically loaded during posttreatment offline analysis on computers not involved in data collection. Participants were blinded to the target of their treatment as neuroreinforcement was performed implicitly; i.e. participants were provided no specific instruction as to what to think about during neuroreinforcement and had no knowledge as to which phobia was being targeted or how many of their multiple phobias would be targeted. Participant strategies were monitored daily to ensure participants had not coincidentally thought about their target (or control) phobia during neuroreinforcement, effectively unblinding themselves. No participants reported thinking about either the target or control phobia during neuroreinforcement.

Decoder construction

Prior to neuroreinforcement, a between-subject machine-learning decoder was trained for the target phobic image category in the ventral temporal (VT) area (Fig. 1). The decoder was constructed using brain data from healthy controls ($N = 22$) using a functional alignment method called hyperalignment.[32] During an initial fMRI session (Fig. 2A), each healthy control viewed the same image data set of 3600 images consisting of 40 categories of animals and objects (e.g. birds, butterflies, snakes, or spiders) (Fig. 1A). Conversely, participants with phobias viewed the same image data set but with their specific phobias removed to avoid unnecessary exposure. Participant-specific decoders were developed using surrogate data based on previous methods,[17] detailed in Supplemental Methods along with task details (Fig. 1B). Importantly, the VT area is a category-selective visual region in which hyperalignment decoding is not affected by fear levels, enabling the use of surrogate brain data from healthy controls to train decoders for phobic participants.[17] Additional research has shown that some ventral visual areas can appear to be predictive of subjective fear ratings.[21] However, these findings likely do not truly reflect "subjective fear" but rather statistical regularities in commonly feared stimuli. This appears to be the case since the same "fear-brain" associations of a fearful person can be detected when using brain data from persons reporting no fear of the stimuli. These findings suggest that this brain signal does not truly represent fear but perhaps statistical regularities of commonly feared animal categories (e.g. spiders). This further suggests that VT representations should be similar among participants regardless of fear levels. Of the potential phobic categories to be selected for treatment for the current participant, the phobia with the highest cross-validated area under the ROC curve scores was blindly selected via computer program as the target for treatment. The within-subject control was also blindly selected through automated random selection from the remaining phobic categories if the participant had more than two phobias. Double-blind target selection was performed in this manner to maximize signal to noise during neuroreinforcement. Importantly, decoder area under the curve during decoder construction should not influence the value of feedback scores during neuroreinforcement. Confirming this, in this study there was no relation between decoder performance during decoder construction and neuroreinforcement scores for the target category ($r(16) = -0.11$, $P = 0.66$; Supplemental Fig. S1A). Moreover, concerning the difference between target and control categories, there was no relation between the difference in decoder performance during decoder construction and the difference in scores calculated during neuroreinforcement ($r(15) = -0.096$, $P = 0.71$; Fig. S1B).

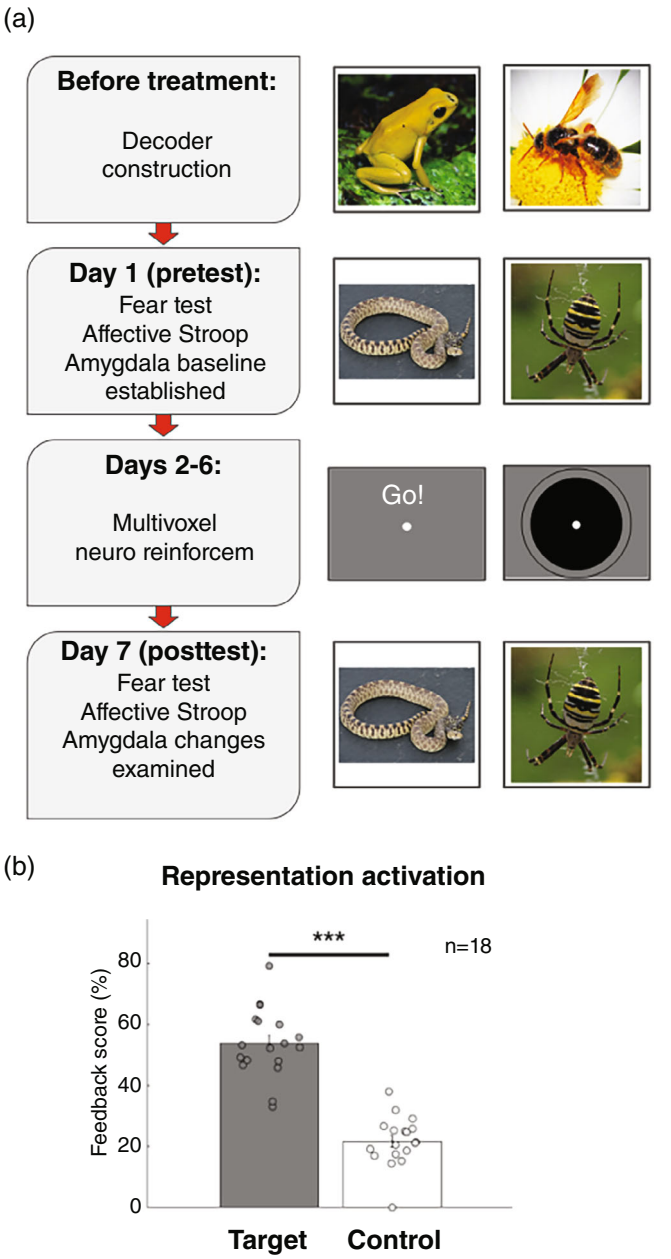

Fig. 2 Study design and activation of target and control representations. (A) Timeline detailing participant activities during each day's functional magnetic resonance imaging session with sample stimuli from each day. Before beginning the treatment program, participants undergo a decoder construction session where they view nonphobic images to enable hyperalignment with healthy controls. On day 1 of treatment, participants complete a pretest in which phobic (and nonphobic) images are rated for fearfulness. Over the next 5 days, participants complete their assigned number of multivoxel neuroreinforcement sessions (1, 3, or 5 days). On day 7, participants complete the same task as a posttest to assess changes in amygdala and skin conductance response to treated and untreated phobias. (B) Representation pattern activation (measured by feedback score) for target phobia compared with control phobia. Target phobia pattern was activated significantly more than control during neuroreinforcement. ***$P < 0.001$.

Preneuroreinforcement and postneuroreinforcement assessments

Each participant with a phobia completed a pretreatment and posttreatment fMRI session (Fig. 2A), during which they completed a fear test as well as an affective Stroop task while their blood oxygen-level–dependent (BOLD) activity was recorded.

Fear test (H1, H2, H3, and H5)

Neural and subjective fear responses were measured to 6-s exposures to photographic images from phobic and neutral animal categories, that included the targeted and control phobias for each participant, following the previous proof-of-concept study.[17] The neutral animal category was randomly selected from animals for which participants self-reported a total absence of fear at screening. Participants reported how fearful each image made them feel on a seven-point Likert scale following each image presentation. See Supplemental Methods for full task details.

Skin conductance response (H2)

SCR recordings were obtained in the fMRI scanner during the fear test. Details of data collection and analysis are reported in Supplemental Methods.

Affective Stroop (H4)

An affective Stroop task assessed reflexive attentional responses to phobic stimuli. Participants made rapid judgments about whether briefly presented animals could fit in their hand. Full task details can be found in Supplemental Methods.

Multivoxel neuroreinforcement

Using multivoxel neuroreinforcement, successful activation of the phobic image category was paired with reward (Fig. 2A). Participants were only aware that the neuroreinforcement task was intended to function as a treatment for phobia. They were blinded to all other information including what the feedback was based on in their brain, how it was calculated, or how many of their phobias this would treat. While participants laid in the fMRI scanner instructed to "use whatever mental strategy they can" to get the best feedback, a neuroreinforcement method[17] was used to reward a nonconsciously represented phobic image category (e.g. spider). Feedback was based on real-time output of the decoder constructed for the individual corresponding to the specific animal phobia selected for targeting. Individual strategies were recorded at the end of each multivoxel neuroreinforcement session. Common strategies reported by participants included thinking about family and friends, memories from the past or plans for the future, imagining oneself doing activities (e.g. exercising, sports, riding rollercoasters), or "mindful" techniques such as focusing on their breath, doing mental math, or simply trying to imagine the feedback circle getting bigger. Participants reported that strategies which seemed to work at one moment did not seem to work later, indicating that one strategy did not work better than others overall.

Each neuroreinforcement run began with an extended rest period of 50 s. Then, an additional rest period of 10 s was collected to determine baseline BOLD activity levels followed by 16 trials of neuroreinforcement. Each trial began with 6 s of rest, followed by 6 s of "induction" where participants modulated their brain activity in an attempt to receive high feedback. Following induction, real-time decoder output was calculated during a 4-s period and then displayed as a green disc for 2 s. This calculation focused on BOLD activity from 4 s after the start of the induction period until 4 s after the induction period ended to account for hemodynamic response delay. The size of the disc directly corresponded to the likelihood estimate such that a 100% likelihood was associated with a maximum disc size (indicated by a visual boundary) and a 0% likelihood was associated with no disc display.

Data processing

Data processing was performed for amygdala response (H1, H4iii, and H5) and affective Stroop (H4) analyses (see Supplemental Methods for data processing details).

Data analysis plan

Amygdala responses were tested with a 2 (condition: target phobia/control phobia) x 2 (time: pretreatment/posttreatment) repeated-measures ANOVA using JASP software (JASP Team 2022). Due to limited sample size (from the COVID-19 pandemic), we were insufficiently powered to analyze neuroreinforcement dosage groups separately, as we had initially preregistered in hypothesis H5. Instead, the neuroreinforcement group (1, 3, or 5 days) was included as a covariate in the ANOVA as was each participant's total number of phobias, as a measure of clinical severity. The between-group data are presented in Figs. S3A, S3B, and S3C and S4A, S4B, and S4C for illustration purposes and the preregistered statistical analysis for H5 is reported in Supplemental Results. To test for a significant reduction in amygdala response for the target phobia category posttreatment compared with the control phobia (H1), we used contrasts of marginal means in JASP to test effects within our ANOVA, which included covariates. These follow-up contrasts were performed on pretreatment and posttreatment activations for the target phobia and control phobia.

Planned *t* tests were performed on pretreatment and posttreatment subjective fear ratings for the target phobia and control phobia, using custom scripts in MATLAB, to test H3. One of the 18 participants was excluded from this analysis of self-reported fear due to not using the button box properly, resulting in 17 participants.

To verify that phobic images were modulating attention as intended, a *t* test was performed on affective Stroop reaction times to phobic images (grouping target and control) and neutral animal images pretreatment (H4.i). For treatment effects (H4.ii), reaction times for correct trials were tested with a 2 (condition: target phobia/control phobia) × 2 (time: pretreatment/posttreatment) repeated-measures ANOVA using JASP software. Dosage group and number of phobias were included as covariates in the model. Similar to the amygdala response analysis, contrasts of marginal means were performed on pretreatment and posttreatment reaction times for the target phobia and control phobia.

Results

A total of 23 participants (Table 1, Table 2, Fig. 3, Table S1) completed pretreatment. Our intended goal of 30 participants, 10 per

Table 1. Participant demographics

Race	Number
White	9
Black	2
Asian/Pacific Islander	9
Other	1
Not reported	2
Ethnicity	
Hispanic	5
Non-Hispanic	18
Sex	
Male	7
Female	16
Nonbinary	0
Age: mean (SD)	26.5 (9.40)
Education level	
High school	4
Some college	3
Associates/2-year degree or higher	16

PCN Psychiatry and Clinical Neurosciences

Table 2. Participant measures (collapsed across randomization groups)

Baseline	Target phobia	Control phobia
ADIS-5 interviewer fear rating (0–8): mean (SD)	5.17 (1.11)	5.91 (1.00)
ADIS-5 interviewer avoidance rating (0–8): mean (SD)	5.30 (1.43)	5.78 (1.24)
Pretreatment	**Target phobia**	**Control phobia**
Fear test amygdala: mean (SD), beta value	0.50 (0.96)	0.31 (0.57)
Fear test subjective fear rating (0–8): mean (SD)	3.63 (1.44)	4.20 (1.12)
Affective Stroop Amygdala: mean (SD), beta value	0.01 (0.31)	0.07 (0.28)
Affective Stroop reaction time: mean (SD), s	1.00 (0.20)	1.01 (0.12)
Posttreatment	**Target phobia**	**Control phobia**
Fear test amygdala: mean (SD), beta value	−0.10 (0.64)	−0.13 (0.86)
Fear test subjective fear rating (0–8): mean (SD)	3.87 (1.40)	4.28 (1.15)
Affective Stroop Amygdala: mean (SD), beta value	−0.05 (0.18)	0.07 (0.29)
Affective Stroop reaction time: mean (SD), s	0.93 (0.16)	0.96 (0.17)

Abbreviation: *ADIS-5*, Anxiety and Related Disorders Interview Schedule for DSM-5.

neuroreinforcement dosage condition, was not achievable due to COVID-19 difficulties.

Double-blind, placebo control and target pattern induction

Following neuroreinforcement, participants were unable to correctly guess the identity of their neuroreinforcement target (see Supplemental Results for more details), indicating they remained blind. Furthermore, the target phobic decoder showed a greater activation likelihood during neuroreinforcement than the control phobic decoder ($t(17)=12.63$, $P < 0.001$) (Fig. 2B, see Supplemental Results and Fig. S2 for more detail), indicating successful nonconscious activation of the target representation during neuroreinforcement.

Amygdala response (H1 and H5)

Before neuroreinforcement, there was a significant amygdala response for both the target phobia ($t(17)=2.20$, $P = 0.042$) and control phobia ($t(17)=2.27$, $P = 0.037$) compared with neutral animals as confirmed by one-sample t tests performed on the baselined parameter estimates. There was no difference in amygdala responses between the target and control phobias prior to neuroreinforcement ($t(17)=0.85$, $P = 0.41$). This indicates successful capturing of threat responding in the amygdala for phobic images.

Following neuroreinforcement, there was a significant interaction between phobia type (target/control) and time (pre/post) shown by a 2 (condition) × 2 (time) repeated-measures ANOVA ($F(1, 15)=5.52$, $P = 0.033$, $\eta_p^2 = 0.269$, Fig. 4A). This result indicates a greater reduction in amygdala response to target phobic images than to control phobic images following neuroreinforcement. After neuroreinforcement, the decrease in amygdala response was significant for the target phobia (contrast of marginal means: $t(25.75) = 2.09$, $P = 0.046$, mean difference [SE] = 0.60 [0.29]; 95% confidence interval (CI), 0.04–1.16) but not the control phobia (contrast of marginal means: $t(25.75) = 1.51$, $P = 0.14$, mean difference [SE] = 0.43 [0.29]; 95% CI, −0.13 to 0.99). These findings support our preregistered hypothesis H1 that amygdala activation would be selectively reduced for the target phobia following neuroreinforcement (see Supplementary Results and Fig. S3 for the H5 results). At posttest, there was no significant difference in amygdala response to the target and control phobias (contrast of marginal means: $t(25.75) = 0.112$, $P = 0.91$).

Skin conductance response (H2)

Our findings did not support hypothesis H2. We did not detect a pretreatment phobia response in SCR data in the nine participants with complete SCR data, using one-sample t tests on baseline-corrected SCR values for either target ($t(8)=0.86$, $P = 0.42$) or control ($t(8)=0.38$, $P = 0.71$) phobias. Given no significant preexisting SCR response, no further statistical testing was performed.

Self-reported fear (H3)

There was no significant change in self-reported fear levels for the 17 participants with complete behavioral data during the fear test in response to either the target phobia ($t(16) = -1.52$, $P = 0.15$, mean difference [SE] = −0.24 [0.16]; 95% CI, −0.86 to 0.13) or the control phobia ($t(16) = -0.56$, $P = 0.58$, mean difference [SE] = −0.08 [0.14]; 95% CI, −0.39 to 0.23), supporting our preregistered hypothesis H3. These findings match previous findings that self-reported fear levels in this task are not modulated by neuroreinforcement.[17]

Affective Stroop (H4)

Results are reported from the 17 participants (five participants in the three-session group, six participants in other groups) with complete brain and behavioral data during the affective Stroop task. Before treatment, reaction times for phobic stimuli were significantly slower compared with responses to neutral stimuli ($t(16)=2.64$, $P = 0.018$, $d = 0.64$, mean difference [SE] = 0.067 [0.025]; 95% CI, 0.017–0.12), confirming our preregistered hypothesis H4i. Slower reaction times for phobic stimuli indicate that attention is successfully captured by phobic stimuli in this task. Importantly, there were no differences between reaction times to target and control phobic images pretreatment ($t(16)=0.91$, $P = 0.38$). Following neuroreinforcement, there was a borderline significant interaction between phobia type (target/control) and time (pre/post) ($F(1, 14)=4.373$, $P = 0.055$; $\eta_p^2 = 0.238$), such that reaction times to the target phobia were faster following neuroreinforcement than they were to the control phobia (Fig. 4B). Importantly, despite not quite reaching the significance threshold, the partial η2 effect size indicates a large effect. Specifically, there were significantly decreased reaction times to target phobia stimuli from pretreatment to posttreatment (contrast of marginal means: $t(18.13) = 2.32$, $P = 0.032$, mean difference [SE] = 0.076 [0.033]; 95% CI, 0.011–0.14) but not for control phobic stimuli (contrast of marginal means: $t(18.13) = 1.30$), $P = 0.21$, mean difference [SE] = 0.043 [0.033]; 95% CI, −0.022 to 0.11. Selectively decreased reaction times for the target phobia indicate that attention is captured less by the target phobia following neuroreinforcement. At the posttest timepoint, there was no significant difference in reaction time to the target phobia compared with the control phobia ($t(18.13) = -1.12$, $P = 0.276$). Reaction time effects by dosage group for H5 are reported in Fig. S4. Amygdala responding during affective Stroop (H4.iii) is reported in Supplemental Results and Supplemental Figs S5A, S5B, S5C, and S5D.

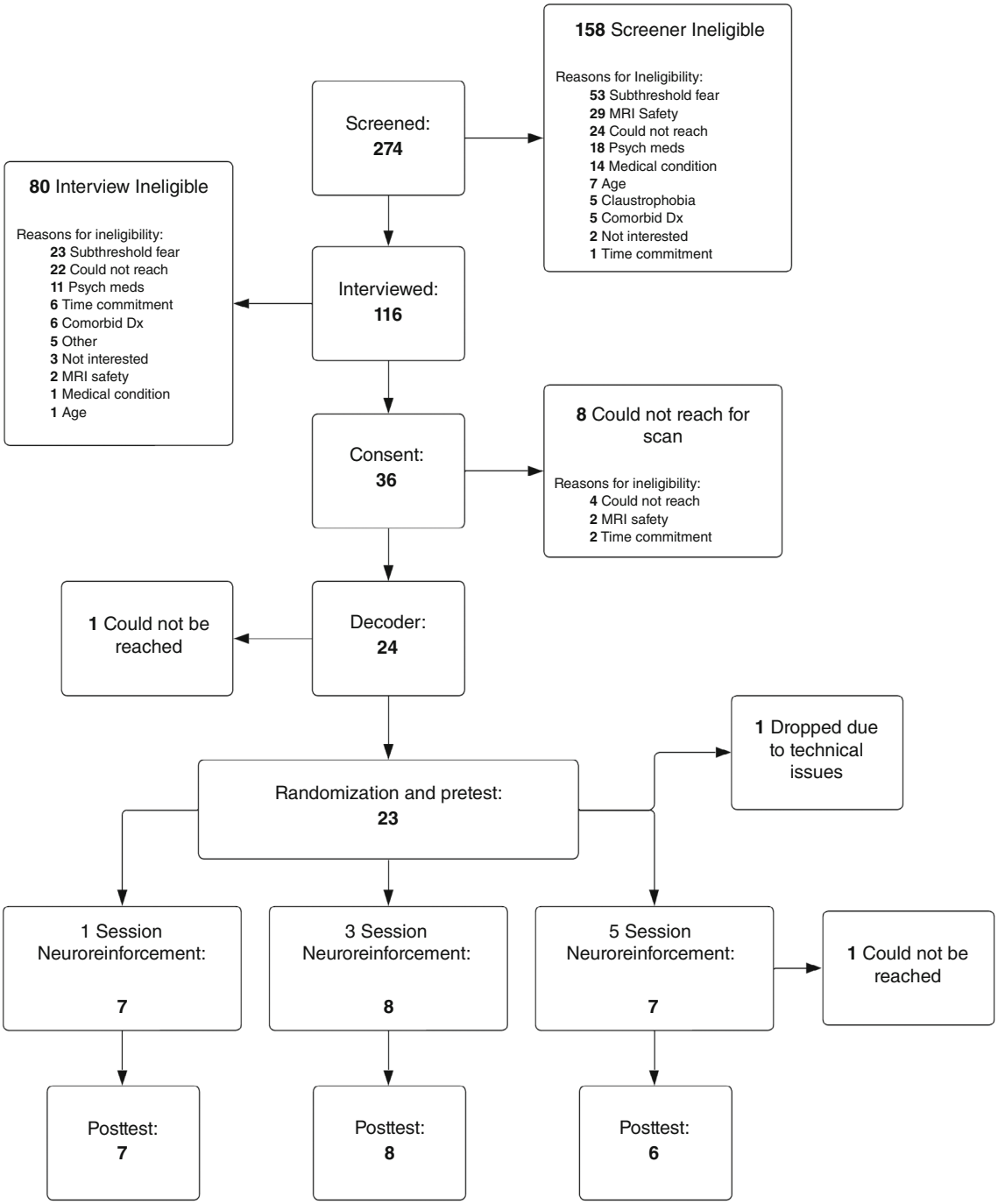

Fig. 3 CONSORT (Consolidated Standards of Reporting Trials) diagram of recruitment flow. Dx, diagnosis; MRI, magnetic resonance imaging.

Discussion

In a double-blind placebo-controlled clinical trial, we investigated whether multivoxel neuroreinforcement could nonconsciously intervene on specific phobia. In line with our hypotheses, we found evidence of large effects for specific reduction in amygdala reactivity (H1) and reduced attentional capture in an affective Stroop task (H4), though the interaction did not reach significance in this latter case. Importantly, these findings were obtained using surrogate participants to determine the target of neuroreinforcement. Consequently, this study supports the ability of decoded neuroreinforcement to be performed without exposing patients to the feared stimulus. Furthermore, our findings were obtained using a double-blind procedure, a level of rigor that is rarely achieved by other psychological interventions.

Decreases in amygdala responses and Stroop reaction times to phobic stimuli represent changes in physiological and reflexive responses to threat.[21,33,34] These changes may represent "preconscious" responses to feared stimuli due to their automatic and reflexive nature.[35–41] Consistent with our hypotheses (H3) and prior findings,[15] no effects were observed with respect to explicit subjective ratings of fear. As subjective fear ratings were not close to ceiling, this may be attributable to the task not eliciting large amounts of fear because participants were only viewing images of animals. Future studies should examine subjective fear to more ecologically valid

PCN Psychiatry and Clinical Neurosciences

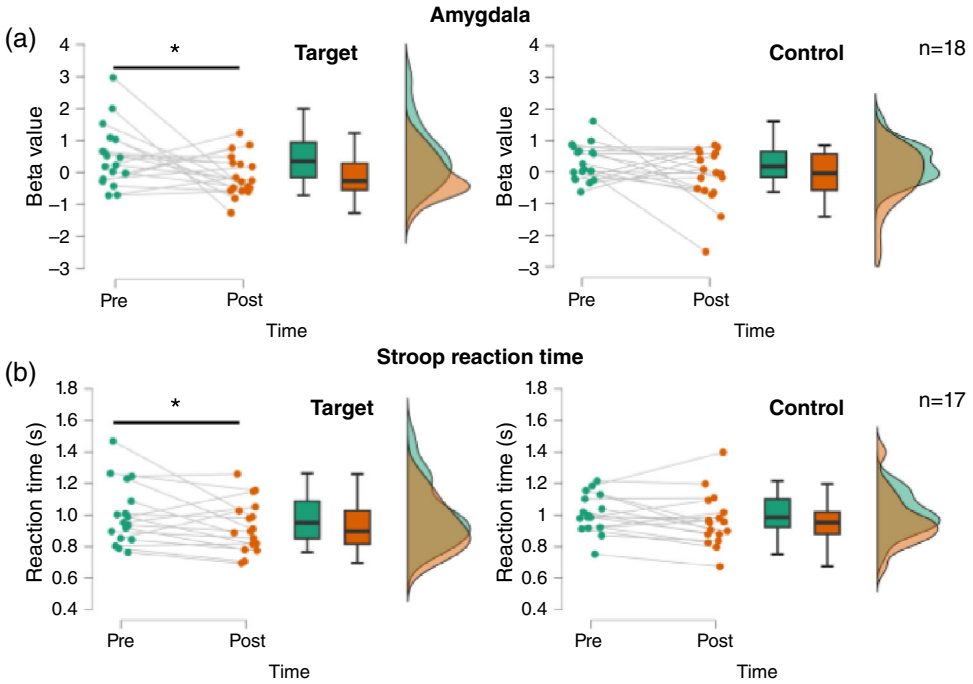

Fig. 4 Changes in fear test amygdala responses and affective Stroop reaction times following neuroreinforcement. (A) Amygdala response in the fear test showed a greater decrease in the target than control phobias following neuroreinforcement. (B) Response times in the affective Stroop task showed a greater decrease in the target than control phobias following neuroreinforcement. *P values < 0.05 indicate significant time (pre/post) effect in the target condition when controlling for days of neuroreinforcement and number of phobias.

in vivo exposures as a function of neuroreinforcement. Alternatively, this pattern of results may suggest that implicit neuroreinforcement is more effective for automatic physiological responses to threat compared with the subjective experience of fear itself. The discordance across response modalities would be consistent with a higher-order theory of emotion in which subjective mental experience operates via different mechanisms than physiological threat responses.[21,33,34,42–45] While an effective treatment would ultimately aim to reduce subjective fear experiences when confronting phobic stimuli, neuroreinforcement could represent an important first step prior to exposure therapy. For example, with reduced physiological threat responses, the reduction of subjective discomfort during traditional exposures may occur at an increased rate as subjective feelings come into alignment with already decreased physiological responding.

Furthermore, following neuroreinforcement, results from the affective Stroop task showed statistical trends with large effects for reaction times decreasing for the target phobia relative to the control phobia (H4ii). In addition to providing further support for specific target engagement by neuroreinforcement, this result suggests that individuals may be less reflexively avoidant of their phobia following neuroreinforcement. If this is the case, patients may be more willing to persist in exposure that is conducted following neuroreinforcement, leading to lower rates of attrition.

To test this hypothesis, future studies should complement neuroreinforcement with a behavioral-approach task to investigate whether physiological symptoms are decreased when approaching the target phobia following neuroreinforcement. If patients are more willing to approach the feared animal following neuroreinforcement, then neuroreinforcement may be a helpful complementary treatment alongside traditional exposure for ensuring the most comfortable treatment regimen possible.

Results of this experiment paired with our previous investigations[9,16,17] have collectively demonstrated the reduction of amygdala activity, implicit behavioral responses, and SCR using multivoxel neuroreinforcement, indicating that extinction learning can occur nonconsciously. Exactly how this is accomplished remains an open question. The rationale for our methodology is based on exposure treatment with preliminary models supporting an exposure-like effect.[9] The pattern of brain activity participants activate during neuroreinforcement corresponds to a category-level visual representation of phobic animals. Our previous study also indicates that activation of this pattern alone does not generate an amygdala response,[17] suggesting that the amygdala response profile is not altered itself during neuroreinforcement. By inducing repeated activation of the visual representation, perhaps local connectivity changes impact how threat and fear systems respond to the natural presence of this representation during perception, achieving an extinction-like effect. Future studies should examine resting-state connectivity patterns before and after neuroreinforcement to determine whether neural changes are localized to visual or emotional processing areas.

The finding that something like extinction can occur nonconsciously using multivoxel neuroreinforcement is consistent with other studies using very brief exposure.[18,23,24] Our results theoretically support this paradigm while satisfactorily eliminating any doubt that some level of conscious awareness or threat processing is responsible for the observed effects in very brief exposure. Future research should explore the overlap and differences between these strategies.

The current study is not without limitations, however. As we did not detect SCR to phobic stimuli in our group of participants, we were unable to test one of our preregistered hypotheses (H2) that neuroreinforcement would lead to reduced phobic SCR responding. This may have arisen from technical limitations, a large portion of participants being nonresponders, or our relatively limited sample size. Similarly, the current study lacked the statistical power to test one of our other main hypotheses (H5): a between-subjects analysis of how much neuroreinforcement is sufficient to achieve the desired outcomes. This limitation is directly attributable to our smaller-than-planned sample size (18 compared with 30 participants), a shortcoming that was caused by the COVID-19 pandemic and funding body policies out of our control. This smaller sample size similarly limits the certainty that can be placed in our observed effect sizes—a limitation that should be addressed in a future study with a larger sample size. Moreover, the current study was limited by lack of a follow-up visit to determine how long effects of neuroreinforcement may last. Future studies should explore how long neuroreinforcement effects last following the intervention by retesting participants weeks or months after neuroreinforcement is completed.

In addition, our primary measures, while important measures for neurocognitive understanding of fear and threat processing, are not traditionally treatment-targeted symptoms of specific phobia. In future

studies, an independent session should be conducted following neuroreinforcement examining real-life avoidance behaviors and fear levels. Future studies will be needed to understand how these observed changes translate to real-life avoidance and fear responses or perhaps using virtual reality as a first step.

While our use of a within-subjects placebo control enabled us to test our intervention in a double-blind fashion, this design did not allow a between-subjects comparison with a placebo group. This should be addressed in a future study using a between-subjects design with treatment and placebo groups based on random assignment.

In summary, this study represents the first clinical trial of multivoxel neuroreinforcement for reducing fear and threat responses in specific phobia. This procedure demonstrated the ability to lessen physiological, reflexive responses to specific phobia through reduced amygdala activation as well as less attentional capture by phobic stimuli. These findings provide a promising foundation to attempt larger-scale replications in clinical cohorts. Through advents in virtual reality, these responses can also be investigated in future studies using more realistic and immersive stimuli.[46–50] This nonconscious procedure produces minimal discomfort in patients with very low rates of attrition. Consequently, neuroreinforcement may serve to complement current conventional psychotherapy approaches while providing a more tolerable experience for patients seeking treatment.

Acknowledgments

The authors would like to thank study coordinators Ana Costello, Annelise Murillo, and Shawn Wang for their assistance to this study through participant recruitment and data collection. H.L. and M.G.C. received financial support from the US National Institute of Mental Health (R61MH113772). H.L. received additional support from Templeton World Charity Foundation (RA537-01). M.K. was supported by AMED under grant number JP18dm0307008. V.T-D. received financial support from the Fonds de Recherche du Québec-Santé (FRQS) and the Fondation de l'Institut Universitaire en Santé Mentale de Montréal. This work was (partially) supported by Innovative Science and Technology for Security grant number JPJ004596, ATLA, Japan.

Disclosure statement

M.K. is an inventor of patents owned by the Advanced Telecommunications Research Institute International related to the present work (PCT/JP2012/078136 [WO2013/068719517] and PCT/JP2014/61543 [WO2014/178322]). C.A.C., H.L., M.G.C., and V.T-D. have no conflicts to declare.

Author contributions

Study design & conception – C.A.C., H.L., M.K., M.G.C., V.T-D. Data acquisition and analysis – C.A.C., V.T-D. Writing & Editing – C.A.C., H.L., M.K., M.G.C., V.T-D.

References

1. 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.
2. Loerinc AG, Meuret AE, Twohig MP, Rosenfield D, Bluett EJ, Craske MG. Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clin. Psychol. Rev.* 2015; **42**: 72–82.
3. Zayfert C, DeViva JC, Becker CB, Pike JL, Gillock KL, Hayes SA. Exposure utilization and completion of cognitive behavioral therapy for PTSD in a "real world" clinical practice. *J. Trauma. Stress* 2005; **18**: 637–645.
4. Eftekhari A, Ruzek JI, Crowley JJ, Rosen CS, Greenbaum MA, Karlin BE. Effectiveness of national implementation of prolonged exposure therapy in veterans affairs care. *JAMA Psychiatry* 2013; **70**: 949–955.
5. Powers MB, Emmelkamp PMG. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J. Anxiety Disord.* 2008; **22**: 561–569.
6. Young KD, Zotev V, Phillips R *et al*. Real-time fMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PLoS One* 2014; **9**: e88785.
7. Gerin MI, Fichtenholtz H, Roy A *et al*. Real-time fMRI neurofeedback with war veterans with chronic PTSD: A feasibility study. *Front. Psychiatry* 2016; **7**: 111.
8. Scheinost D, Stoica T, Saksa J *et al*. Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. *Transl. Psychiatry* 2013; **3**: e250.
9. Chiba T, Kanazawa T, Koizumi A *et al*. Current status of neurofeedback for post-traumatic stress disorder: A systematic review and the possibility of decoded neurofeedback. *Front. Hum. Neurosci.* 2019; **13**: 233.
10. Stoeckel LE, Garrison KA, Ghosh SS *et al*. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *Neuroimage Clin.* 2014; **5**: 245–255.
11. Mennen AC, Turk-Browne NB, Wallace G *et al*. Cloud-based functional magnetic resonance imaging neurofeedback to reduce the negative attentional bias in depression: A proof-of-concept study. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2021; **6**: 490–497.
12. Young KD, Siegle GJ, Zotev V *et al*. Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: Effects on symptoms and autobiographical memory recall. *Am. J. Psychiatry* 2017; **174**: 748–755.
13. Cohen JD, Daw N, Engelhardt B *et al*. Computational approaches to fMRI analysis. *Nat. Neurosci.* 2017; **20**: 304–313.
14. Watanabe T, Sasaki Y, Shibata K, Kawato M. Advances in fMRI real-time neurofeedback. *Trends Cogn. Sci.* 2017; **21**: 997–1010.
15. deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* 2015; **18**: 470–475.
16. Koizumi A, Amano K, Cortese A *et al*. Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nat. Hum. Behav.* 2016; **1**: 6.
17. Taschereau-Dumouchel V, Cortese A, Chiba T, Knotts JD, Kawato M, Lau H. Towards an unconscious neural reinforcement intervention for common fears. *Proc. Natl. Acad. Sci. USA* 2018; **115**: 3470–3475.
18. Siegel P, Cohen B, Warren R. Nothing to fear but fear itself: A mechanistic test of unconscious exposure. *Biol. Psychiatry* 2022; **91**: 294–302.
19. Taschereau-Dumouchel V, Cushing CA, Lau H. Real-time functional MRI in the treatment of mental health disorders. *Annu. Rev. Clin. Psychol.* 2022; **18**: 125–154.
20. Siegel P, Wang Z, Murray L *et al*. Brain-based mediation of nonconscious reduction of phobic avoidance in young women during functional MRI: A randomised controlled experiment. *Lancet Psychiatry* 2020; **7**: 971–981.
21. Taschereau-Dumouchel V, Kawato M, Lau H. Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Mol. Psychiatry* 2020; **25**: 2342–2354.
22. Siegel P, Warren R, Wang Z *et al*. Less is more: Neural activity during very brief and clearly visible exposure to phobic stimuli. *Hum. Brain Mapp.* 2017; **38**: 2466–2481.
23. Siegel P, Warren R. Less is still more: Maintenance of the very brief exposure effect 1 year later. *Emotion* 2013; **13**: 338–344.
24. Siegel P, Warren R. The effect of very brief exposure on experienced fear after in vivo exposure. *Cogn. Emot.* 2013; **27**: 1013–1022.
25. Rance M, Walsh C, Sukhodolsky DG *et al*. Time course of clinical change following neurofeedback. *Neuroimage* 2018; **181**: 807–813.
26. Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron* 2004; **43**: 897–905.
27. Delgado MR, Nearing KI, LeDoux JE, Phelps EA. Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron* 2008; **59**: 829–838.
28. LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA. Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron* 1998; **20**: 937–945.
29. Wen Z, Raio CM, Pace-Schott EF *et al*. Temporally and anatomically specific contributions of the human amygdala to threat and safety learning. *Proc. Natl. Acad. Sci. USA* 2022; **119**: e2204066119.
30. Lissek S, Powers AS, McClure EB *et al*. Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behav. Res. Ther.* 2005; **43**: 1391–1424.
31. Brown T, Barlow D. *Anxiety and Related Disorders Interview Schedule for DSM-5 (ADIS-5)® - Adult Version: Client Interview Schedule 5-Copy Set*. Oxford University Press, Oxford, NY, 2014; 84 (Treatments That Work).

32. Haxby JV, Guntupalli JS, Connolly AC *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 2011; **72**: 404–416.
33. LeDoux JE, Pine DS. Using neuroscience to help understand fear and anxiety: A two-system framework. *Am. J. Psychiatry* 2016; **173**: 1083–1093.
34. Taschereau-Dumouchel V, Michel M, Lau H, Hofmann SG, LeDoux JE. Putting the "mental" back in "mental disorders": A perspective from research on fear and anxiety. *Mol. Psychiatry* 2022; **27**: 1322–1330.
35. Mobbs D, Marchant JL, Hassabis D *et al*. From threat to fear: The neural organization of defensive fear systems in humans. *J. Neurosci.* 2009; **29**: 12236–12243.
36. Larson CL, Schaefer HS, Siegle GJ, Jackson CAB, Anderle MJ, Davidson RJ. Fear is fast in phobic individuals: Amygdala activation in response to fear-relevant stimuli. *Biol. Psychiatry* 2006; **60**: 410–417.
37. McFadyen J, Mermillod M, Mattingley JB, Halász V, Garrido MI. A rapid subcortical amygdala route for faces irrespective of spatial frequency and emotion. *J. Neurosci.* 2017; **37**: 3864–3874.
38. Méndez-bértolo C, Moratti S, Toledano R *et al*. A fast pathway for fear in human amygdala. *Nat. Neurosci.* 2016; **19**: 1041–1049.
39. Cushing CA, Im HY, Adams RB, Ward N, Kveraga K. Magnocellular and parvocellular pathway contributions to facial threat cue processing. *Soc. Cogn. Affect. Neurosci.* 2019; **14**: 151–162.
40. Cushing CA, Im HY, Adams RB *et al*. Neurodynamics and connectivity during facial fear perception: The role of threat exposure and signal congruity. *Sci. Rep.* 2018; **8**: 2776.
41. Adams RB, Franklin RG, Kveraga K *et al*. Amygdala responses to averted vs direct gaze fear vary as a function of presentation speed. *Soc. Cogn. Affect. Neurosci.* 2012; **7**: 568–577.
42. Brown R, Lau H, LeDoux JE. Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* 2019; **23**: 754–768.
43. LeDoux JE, Lau H. Seeing consciousness through the lens of memory. *Curr. Biol.* 2020; **30**: R1018–R1022.
44. LeDoux JE. Thoughtful feelings. *Curr. Biol.* 2020; **30**: R619–R623.
45. LeDoux J. *Anxious: Using the Brain to Understand and Treat Fear and Anxiety.* Penguin, New York, 2016; 482.
46. Morina N, Ijntema H, Meyerbröker K, Emmelkamp PMG. Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behav. Res. Ther.* 2015; **74**: 18–24.
47. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 2011; **12**: 752–762.
48. Dunsmoor JE, Ahs F, Zielinski DJ, LaBar KS. Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. *Neurobiol. Learn. Mem.* 2014; **113**: 157–164.
49. Shiban Y, Reichenberger J, Neumann ID, Mühlberger A. Social conditioning and extinction paradigm: A translational study in virtual reality. *Front. Psychol.* 2015; **6**: 400.
50. Tröger C, Ewald H, Glotzbach E, Pauli P, Mühlberger A. Does pre-exposure inhibit fear context conditioning? A virtual reality study. *J. Neural Transm.* 2012; **119**: 709–719.

Supporting Information

Additional supporting information can be found online in the Supporting Information section at the end of this article.