Psychological Medicine

cambridge.org/psm

Decoding neural reactivation of threat during fear learning, extinction, and recall in a randomized clinical trial of L-DOPA among women with PTSD

Josh M. Cisler[1,2], Joseph E. Dunsmoor[1,2], Anthony A. Privratsky[3] and G. Andrew James[4]

[1]Department of Psychiatry and Behavioral Sciences, Dell Medical School, University of Texas at Austin, Austin, TX, USA; [2]Institute for Early Life Adversity Research, Dell Medical School, University of Texas at Austin, Austin, TX, USA; [3]University of Utah Health, Salt Lake City, UT, USA and [4]Brain Imaging Research Center, Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Original Article

Cite this article: Cisler JM, Dunsmoor JE, Privratsky AA, James GA (2024). Decoding neural reactivation of threat during fear learning, extinction, and recall in a randomized clinical trial of L-DOPA among women with PTSD. *Psychological Medicine* **54**, 1091–1101. https://doi.org/10.1017/S0033291723002891

Received: 21 June 2023
Revised: 21 August 2023
Accepted: 7 September 2023
First published online: 9 October 2023

Keywords:
dopamine; fear extinction; multivariate pattern analysis; PTSD

Corresponding author:
Josh M. Cisler;
Email: josh.cisler@austin.utexas.edu

Abstract

Background. Laboratory paradigms are widely used to study fear learning in posttraumatic stress disorder (PTSD). Recent basic science models demonstrate that, during fear learning, patterns of activity in large neuronal ensembles for the conditioned stimuli (CS) begin to reinstate neural activity patterns for the unconditioned stimuli (US), suggesting a direct way of quantifying fear memory strength for the CS. Here, we translate this concept to human neuroimaging and test the impact of post-learning dopaminergic neurotransmission on fear memory strength during fear acquisition, extinction, and recall among women with PTSD in a re-analysis of previously reported data.

Methods. Participants ($N = 79$) completed a context-dependent fear acquisition and extinction task on day 1 and extinction recall tests 24 h later. We decoded activity patterns in large-scale functional networks for the US, then applied this decoder to activity patterns toward the CS on day 1 and day 2.

Results. US decoder output for the CS+ increased during acquisition and decreased during extinction in networks traditionally implicated in human fear learning. The strength of US neural reactivation also predicted individuals skin conductance responses. Participants randomized to receive L-DOPA ($n = 43$) following extinction on day 1 demonstrated less US neural reactivation on day 2 relative to the placebo group ($n = 28$).

Conclusion. These results support neural reactivation as a measure of memory strength between competing memories of threat and safety and further demonstrate the role of dopaminergic neurotransmission in the consolidation of fear extinction memories.

Posttraumatic stress disorder (PTSD) is characterized by intrusive recollections of the traumatic event, avoidance of stimuli related to the trauma, negative changes in mood and cognition, and hyperarousal (Weathers, Keane, & Davidson, 2001). Existing evidence-based treatments for PTSD strongly rely on exposure-based modalities, which are hypothesized to work via the basic mechanisms of classical conditioning and extinction (Foa, Rothbaum, Riggs, & Murdock, 1991; Rothbaum & Davis, 2003). Accordingly, laboratory paradigms of fear conditioning and extinction are widely used to test and develop novel modalities to enhance exposure-based therapies (Crombie et al., 2023, 2021; Dunsmoor, Cisler, Fonzo, Creech, & Nemeroff, 2022; Raij et al., 2018).

Recent animal and human laboratory work suggests a key role for dopamine signaling in the consolidation of fear extinction memories (Kalisch, Gerlicher, & Duvarci, 2019). Animal models using pharmacological and chemogenetic manipulations demonstrate a critical role for dopaminergic neurons projecting to the striatum during consolidation of fear extinction memories (Bouchet et al., 2017; Haaker et al., 2013; Luo et al., 2018). Human studies among healthy men similarly demonstrate that L-DOPA, a dopamine precursor that broadly increases dopamine signaling, delivered after fear extinction learning results in less return of fear 24 h later (Gerlicher, Tüscher, & Kalisch, 2018; Haaker et al., 2013). In a sample of women with PTSD, we previously demonstrated that post-extinction L-DOPA decreased skin conductance responses (SCRs) and anterior insula activation to conditioned stimuli (CS) following a fear reinstatement procedure (Cisler et al., 2020). Here it is important to differentiate *learning* the extinction memory *v. consolidating* the extinction memory, as this body of data supports manipulations of dopaminergic signaling more specifically during the consolidation window as a means to enhance the consolidation of fear extinction memory, thereby increasing likelihood of fear extinction memory retrieval upon subsequent CS presentation.

Testing the impact of laboratory manipulations, such as dopamine signaling, on fear extinction consolidation success necessitates laboratory operational definitions of fear memory strength (i.e. since extinction memory strength is inferred indirectly through degree of fear memory strength). In humans, common measures to evaluate the success of extinction learning include psychophysiological responding (e.g. skin conductance) and functional MRI activation (e.g. amygdala activity) (Bach & Melinscak, 2020; Fullana et al., 2016, 2018; Ojala & Bach, 2020) to CS at an extinction recall test some time following extinction memory formation. The underlying conceptualization is that presentation of an extinguished conditioned stimulus elicits competing retrieval of the fear memory and the extinction memory (Bouton, 2004), with the relative degree of retrieval between these memories thereby determining the degree of measurable fear responding (e.g. the magnitude of autonomic arousal). In this way, downstream measures of fear provide an indirect index of fear memory strength (*v.* extinction memory strength). Recent neurophysiological animal models offer a novel methodology to quantify fear memory strength more directly through populations of neural activity (Herry & Jercog, 2022). Longitudinal monitoring of dynamics within large populations of amygdala neurons demonstrates that, during fear conditioning, the neural ensemble representation of the conditioned stimulus (CS) reshapes to resemble the representation of the unconditioned stimulus (US) (Grewe et al., 2017; Zaki et al., 2022; Zhang & Li, 2018). Beyond suggesting a supervised learning model by which fear learning operates (Knudsen, 1994), these data also suggest a direct and intuitive means of quantifying fear memory strength: the degree to which the CS reactivates the representation of the threatening outcome it had predicted in the past, i.e. the US (Grewe et al., 2017). This type of direct measure of fear memory strength, if translated to humans, could further bolster support for laboratory manipulations hypothesized to enhance fear extinction learning and decrease fear memory strength.

Multivariate pattern analysis is increasingly used in human fear conditioning and extinction fMRI studies (Hennings, Cooper, Lewis-Peacock, & Dunsmoor, 2022), and is an appropriate statistical approach for testing CS reactivation of the US during learning in humans. A decoder trained to discriminate activity patterns for the US itself can be applied to activity patterns for the CS during conditioning and extinction, allowing CS elicited reactivation of the US to be quantified. Here, we use MVPA to translate the hypothesis that a CS is sufficient to reactivate neural activity patterns of a US (mild electrical shock) that it has been associated with in the past. We leverage a previously reported multi-day fear conditioning/extinction and extinction recall data set in women with PTSD who received L-DOPA or placebo following extinction on day 1, which allows us to test the degree of US neural reactivation 24 h following extinction learning that has putatively been enhanced through consolidation-dependent dopaminergic signaling (Cisler et al., 2020). This allows us to assess whether this measure of fear memory retrieval is effectively diminished by manipulating the strength of the extinction memory representation through increased dopamine signaling during the consolidation window. Based on meta-analytic results for neural encoding of fear conditioning and pain in humans (Biggs et al., 2020; Fullana et al., 2016; Palermo, Benedetti, Costa, & Amanzio, 2015; Xu et al., 2020), we focus on US representations in functional networks consisting of anterior insula, dorsal anterior cingulate, striatum, and superior temporal gyrus. Given the role of neuronal ensembles in the amygdala in animal models of CS reactivation of US representations, we include an additional bilateral mask of amygdala (Grewe et al., 2017; Zhang & Li, 2018). We also test a whole-brain mask of grey matter (GM) voxels to identify a brain-wide US representation unconstrained within any specific network. We hypothesize that (1) CS reactivation of US representations increases during fear acquisition and decreases during fear extinction, (2) CS reactivation of US representations predicts downstream psychophysiological responding and self-reported US expectancy, and (3) participants receiving L-DOPA after day 1 demonstrate decreased CS reactivation of US representations (i.e. decreased fear memory strength) on day 2.

Method

Participants, assessments, and inclusion criteria

Participants consisted of 79 women with PTSD related to interpersonal violence with usable fMRI data (from an initial randomized sample of 91; see Supplementary material for further inclusion and exclusionary criteria). Table 1 lists demographic and clinical characteristics of the sample. All procedures were approved by the appropriate IRBs and all patients provided informed consent.

Double-blind randomization

Participants were randomized using blocked stratified randomization (see Supplementary material). Online Supplementary Fig. S1 provides participant enrollment and attrition throughout the design.

Fear conditioning, fear extinction, and fear recall task

The task used here (online Supplementary Fig. S2) was modeled after a prior study among healthy adults testing the impact of L-DOPA on context renewal (Haaker et al., 2013). The US was an electric shock. CS consisted of triangles and circles. The US occurred 2.5 s following CS+ onset with a 50% reinforcement schedule during the acquisition phase. Colored backgrounds distinguished the acquisition and extinction contexts. No shocks occurred during the extinction phase. The task alternated between acquisition and extinction phases, with two presentations of each phase. US expectancy ratings were collected three times in each block on day 1. See the Supplementary material for more details.

Participants completed the day 2 fear recall task 24 h following day 1 learning. The task presented two CS+ and CS− stimuli per context (acquisition and extinction contexts from day 1) for a total of three context presentations (six total CS presentations each), with no shock presentations. During this initial recall test (i.e. initial fear recall), responding to the CSs in the extinction context reflects extinction recall, whereas responding to the CSs in the acquisition context reflects renewal. After this initial fear recall test, participants then received a single unsignaled US presentation to promote fear reinstatement before completing the recall task again (i.e. reinstatement). In this task, extinction retrieval, context renewal, and fear reinstatement are respectively operationalized on day 2 through fear responding in the extinction context, when returning to the acquisition context, and following the US reinstatement procedure.

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

Table 1. Comparison of clinical and demographic characteristics between participant groups (placebo *v.* 100 mg *v.* 200 mg)

Variable	Placebo (a) *N* = 34	100 mg (b) *N* = 28	200 mg (c) *N* = 29	*p* values
Age (year)	33.8 (8.8)	34.8 (9.7)	34.5 (8.6)	*p*(ab) = 0.655
				p(ac) = 0.734
				p(bc) = 0.901
Education (year)	15.5 (2.9)	14.8 (2.0)	15.0 (2.6)	*p*(ab) = 0.251
				p(ac) = 0.457
				p(bc) = 0.713
Ethnicity				*p*(abc) = 0.168
White (%)	76.5	60.7	82.1	
African America (%)	8.8	28.6	17.2	
Asian	0	0	0	
Native American	2.9	0	0	
Hispanic	5.9	0	3.5	
Pacific Islander	0	0	0	
Other	5.9	10.7	0	
IQ				*p*(ab) = 0.292
				p(ac) = 0.393
Verbal	104.3 (24.0)	98.3 (19.2)	99.4 (20.0)	*p*(bc) = 0.839
Digit span	9.8 (2.6)	9.5 (3.3)	9.7 (1.9)	*p*(ab) = 0.695
				p(ac) = 0.850
				p(bc) = 0.799
Direct assault types (#)	6.0 (2.6)	5.6 (2.9)	5.3 (3.5)	*p*(ab) = 0.580
				p(ac) = 0.347
				p(bc) = 0.697
Sexual assault (%)	97.1	92.9	89.7	*p*(abc) = 0.493
Physical assault (%)	88.2	78.6	75.9	*p*(abc) = 0.411
Physical abuse (%)	64.7	60.7	51.7	*p*(abc) = 0.570
Age first assault (year)	8.9 (6.7)	9.4 (6.5)	9.7 (7.0)	*p*(ab) = 0.735
				p(ac) = 0.616
				p(bc) = 0.870
Age last assault (year)	28.4 (9.7)	29.3 (11.2)	27.9 (10.6)	*p*(ab) = 0.752
				p(ac) = 0.843
				p(bc) = 0.641
Time since last assault (year)	5.3 (6.7)	5.5 (7.7)	6.6 (7.8)	*p*(ab) = 0.908
				p(ac) = 0.491
				p(bc) = 0.611
Current mood disorder (%)	26.5	42.9	34.5	*p*(abc) = 0.399
Current comorbid anxiety disorder (%)	70.6	57.1	75.9	*p*(abc) = 0.294
Current GAD (%)	50.0	39.3	62.1	*p*(abc) = 0.227
CAPS-V total severity	43.8 (11.3)	40.1 (10.2)	42.3 (11.9)	*p*(ab) = 0.191
				p(ac) = 0.615
				p(bc) = 0.464
Time between first day of menstrual cycle and fear ext. paradigm (days)	23.7 (6.7)	24.4 (6.0)	25.8 (15.5)	*p*(ab) = 0.818

(*Continued*)

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

Table 1. (*Continued.*)

Variable	Placebo (a) $N = 34$	100 mg (b) $N = 28$	200 mg (c) $N = 29$	p values
				p(ac) = 0.708
				p(bc) = 0.828
Birth control (%)	50.0	53.6	48.3	p(abc) = 0.920
Estradiol concentration[a] (pg/mL)	1.45 (0.82)	1.43 (0.71)	1.34 (0.49)	p(ab) = 0.938
				p(bc) = 0.695
				p(ac) = 0.659
Daily cigarette smoker	17.6	25.0	17.2	p(abc) = 0.706

GAD, generalized anxiety disorder; CAPS, Clinician Administered PTSD Scale DSM-V.
Note: Verbal IQ was assessed from the Receptive One-Word Picture Vocabulary Test. Digit span is from the Wechsler Adult Intelligence Scale-IV.
[a]Estradiol concentration was calculated using enzyme immunoassay upon samples collected immediately following the second scan session. Salivary samples were only available among a subset of participants across both sites and drug groups; $N_a = 21$, $N_b = 18$, $N_c = 15$. Usable imaging data available for $n = 28$ placebo, $n = 19$ 100 mg, and $n = 24$ L-DOPA.

Study medication

Among the participants with usable fMRI data utilized for this analysis, $n = 28$ were randomized with a double blind to placebo, $n = 19$ randomized to 100/25 mg L-DOPA/carbidopa, and $n = 24$ randomized to 200/50 mg L-DOPA/carbidopa. While the study was designed to probe possible dose-dependency effects, our prior analysis suggested generally comparable effects of both active doses on fear responding following reinstatement (Cisler et al., 2020). Thus, to conserve power, the L-DOPA dose groups are combined for this analysis.

Independent component analysis

See the Supplementary material for details regarding the independent component analysis (Calhoun, Adali, Pearlson, & Pekar, 2001). Five components were deemed functional networks of interest for the current analysis, based on prior meta-analyses of fear conditioning and pain processing (Biggs et al., 2020; Fullana et al., 2016, 2018; Palermo et al., 2015; Xu et al., 2020). One of the ICA networks included bilateral nucleus accumbens and dorsal regions of bilateral amygdala. To ensure adequate coverage of amygdala, we created an additional mask of bilateral amygdala defined from the Harvard-Oxford Atlas. We additionally included a brain-wide mask constrained within GM voxels to define a whole-brain GM US reactivation model.

Multivariate pattern analyses

Figure 1 provides an overview of the analytical approach, which is in direct accord with our previous MVPA investigations (Cisler, Tamman, & Fonzo, 2023; Moughrabi et al., 2022).

Decoding the US

The first step was to decode network activity patterns for the US. For this step, each participants' trial-by-trial activation patterns at the time of the US and at the comparable time of non-reinforced CS+ during the acquisition phase were characterized using a least-squares approach (Mumford, Turner, Ashby, & Poldrack, 2012) across all trials (i.e. AFNI's 3dLSS). All trials here are for the CS+ during acquisition, allowing the decoder to isolate the US specifically. The resulting timepoint × voxel matrices were centered within each timepoint to ensure no differences in overall activation across trials. Support vector machines (SVM, with radial basis function kernel and cost parameter $C = 1$) implemented in Matlab through libsvm (Chang & Lin, 2011), were used to decode US *v.* no-US (binary classification). We established the accuracy of the decoders using fivefold cross-validation (i.e. the sample was split into five folds of participants, with one set held out as the test set and repeated until each fold of participants has served as the test set), with 10 separate iterations to account for any variability in partitioning into the folds. The US decoder accuracy was defined as the mean of sensitivity and specificity.

CS reactivation of US representations during day 1

After testing accuracy of the US decoder, the next step was to apply the decoders to participants' data during each CS presentation during each context of the task. 3dLSS was used to define trial-by-trial activation for each CS across the task, providing unique β coefficients for each CS+ and CS− during the acquisition and extinction contexts. Only non-reinforced CS+ trials were included in these analyses. Following recent work (Cisler et al., 2023; Moughrabi et al., 2022; Zhou et al., 2021), a leave-one-out approach was used, such that a subject was designated as the left-out test subject, the US decoders were trained on all remaining participants' US *v.* no-US data except for the left-out test subject, and the resulting US decoders were applied to the left-out participant's data for each day 1 CS. This process was repeated for each subject. This resulted in hyperplane distances representing the degree to which each CS in each phase reactivated the US representation. This process was repeated separately for each network of interest.

CS reactivation of US during day 2

A similar approach was used for day 2 responding to each CS during each phase. 3dLSS was used to define trial-by-trial activation for each CS during each context, for both initial recall and following reinstatement. A leave-one-out approach was used, such that a subject was designated as the left-out test subject, the US decoders were trained on all remaining participants' US *v.* no-US data from day 1 except for the left-out test subject, and the resulting US decoders were applied to the left-out participant's data for each day 2 CS.

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

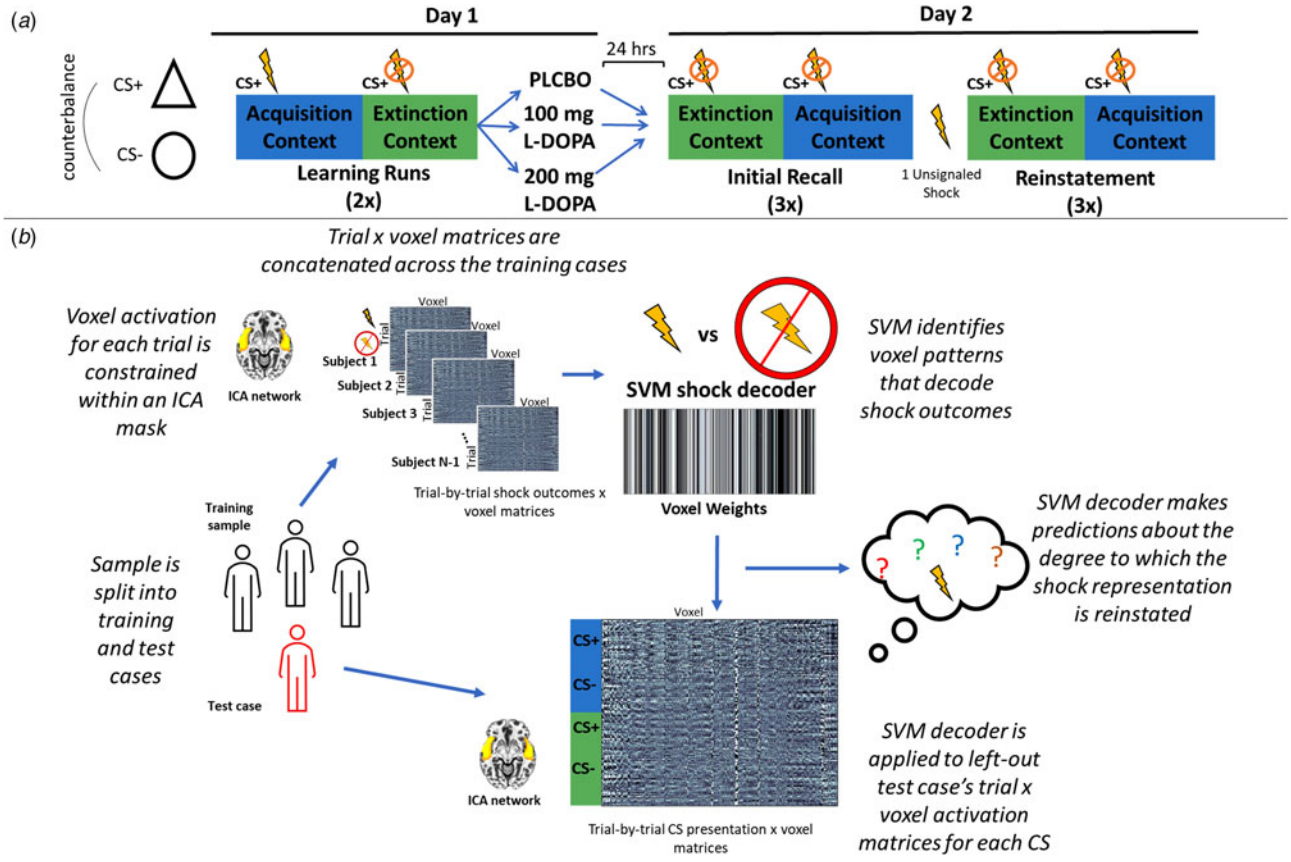

Figure 1. (a) Overview of the randomized clinical trial design. On day 1 participants alternated between a fear acquisition and fear extinction context, detonated by differing colored backgrounds, in which the CS+ (either a triangle or circle, counterbalanced) did or did not predict a mild electric shock, respectively. Participants then received either placebo, 100 mg, or 200 mg of L-DOPA. Participants returned 24 h later for an initial fear recall test, in which they were presented the CS in each context in the absence of any shocks. Participants then underwent a fear reinstatement procedure in which they received a single unsignaled shock, and then underwent the same fear recall task. (b) Overview of the MVPA decoding approach. On day 1, participants were split into a training set and a left out test case. The training sample's trial-by-trial β coefficients, which corresponded to shock delivery *v.* shock absence on CS+ trials in the acquisition context, were constrained within a given ICA network. These trial × voxel matrices were then used to train an MVPA decoder to predict the occurrence of the shock based on patterns of activity in the network. This MVPA decoder was then applied to either day 1 or day 2 for the left-out test case's trial-by-trial β coefficients for each CS in each phase, resulting in a prediction for each CS of the degree to which the US representation was reactivated.

MVPA linear mixed-effects models

We tested hypotheses about the decoder output using linear mixed-effects models (LMEMs; Matlab's fitlme function). For day 1, we modeled trial-by-trial US decoder output (i.e. CS reactivation of US representations) as a function of CS type (CS+ *v.* CS −), context (acquisition *v.* extinction), and group (placebo *v.* L-DOPA) with additional covariates for PTSD symptom severity, education, age, head motion, and site, and including a random intercept for subject. For day 2, given that fear extinction learning continues in the absence of CS+ reinforcement, we focused on the first block (i.e. first two CS presentations) for each context (acquisition and extinction) for initial fear recall. For reinstatement (i.e. after the single isolated US presentation), we compared the last block of the initial fear recall test for each phase (i.e. last two CS+ and CS− in each context prior to reinstatement) and the first block following reinstatement (i.e. first two CS+ and CS− in each context following reinstatement). We corrected for multiple comparisons with Bonferroni correction ($p < 0.05/7$ tests = corrected $p < 0.0071$).

We tested relationships between decoder output (i.e. CS reactivation of US representations) and SCRs with LMEMs that were identical to those described above except for the addition of trial-by-trial SCRs as an interactive factor. We conducted a similar analysis testing relationship of decoder output with US expectancy ratings, which were only available on day 1 (see Supplementary material).

Results

US decoder accuracy

As indicated in Fig. 2, US decoder accuracy was above chance for all networks tested. Decoder accuracy for the bilateral amygdala was notably lower compared to the other networks. Decoder accuracy for the brain-wide GM mask was notably higher compared to the other networks. Further, Fig. 2a provides the voxel feature weights, following transformation to forward encoding (Haufe et al., 2014), from the SVM model for the GM mask, which demonstrates that voxels that contributed strongly to US predictions corresponded to the ICA networks.

CS reactivation of US representations during day 1 fear acquisition and extinction

As indicated in Fig. 3, there were significant CS × context interactions for decoder output for the dACC network, $t(8877) = 3.92$,

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

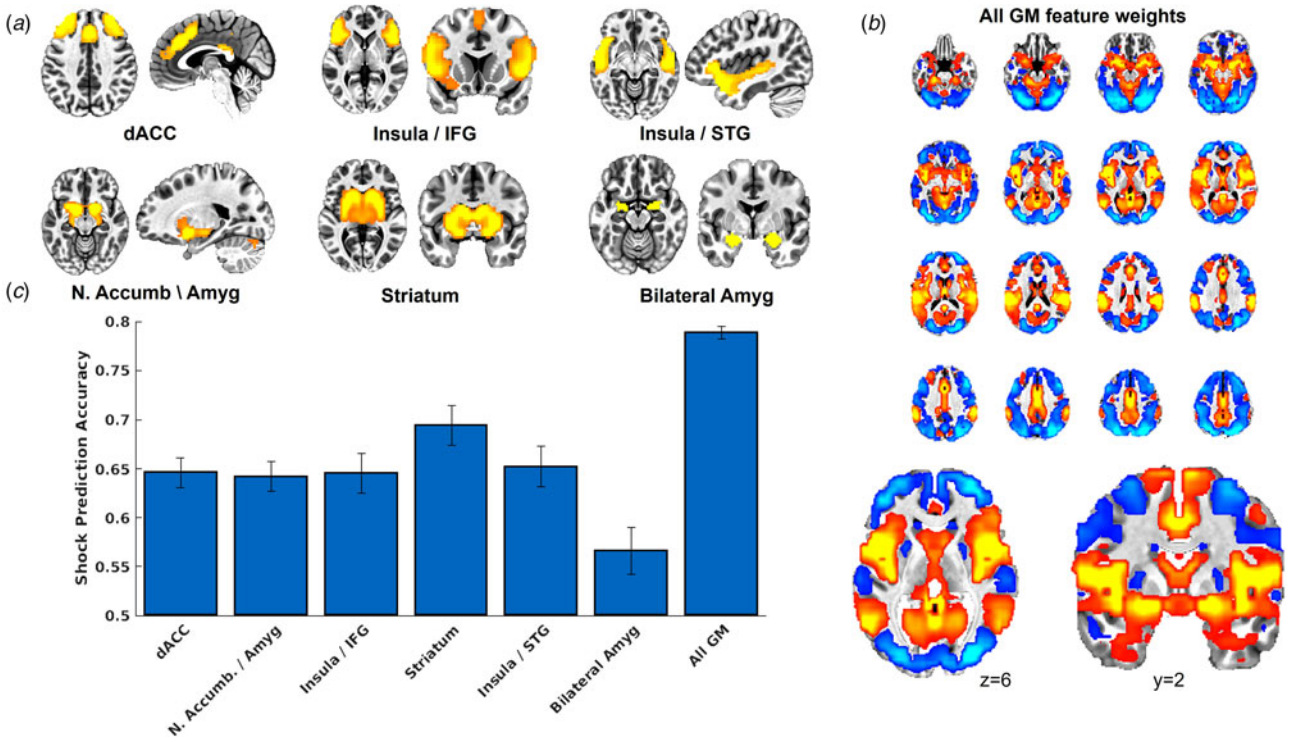

Figure 2. (a) Depiction of the ICA networks tested here. We created an additional mask of complete bilateral amygdala in order to ensure adequate testing of the amygdala's role in representing the US in human neuroimaging data. (b) We also tested a whole-brain grey matter (GM) voxel mask, which demonstrated the highest classification accuracy (see panel c), and we present the voxel feature weights from the SVM model following transformation to forward encoding (Haufe et al., 2014). We also present an expanded image of an axial slice and a coronal slice to demonstrate that the voxel feature weights resemble the ICA networks tested, here demonstrating strong positive weights in regions consistent with the insula/STG network (see panel a) and less strong positive weights in the amygdala. (c) Results from the cross-validation analyses from day 1 that tested the accuracy of the MVPA decoders to predict the occurrence *v.* absence of the shock. Error bars represent the range of the fivefold cross-validation accuracy across 10 iterations.

$p < 0.001$, and insula/STG network $t(8877) = 4.00$, $p < 0.001$. No significant US reactivation was observed for the other networks controlling for multiple corrections.

To further understand the temporal dynamics of CS reactivation of US representations, we divided the task into six blocks (early, middle, and late for each run) and included an additional

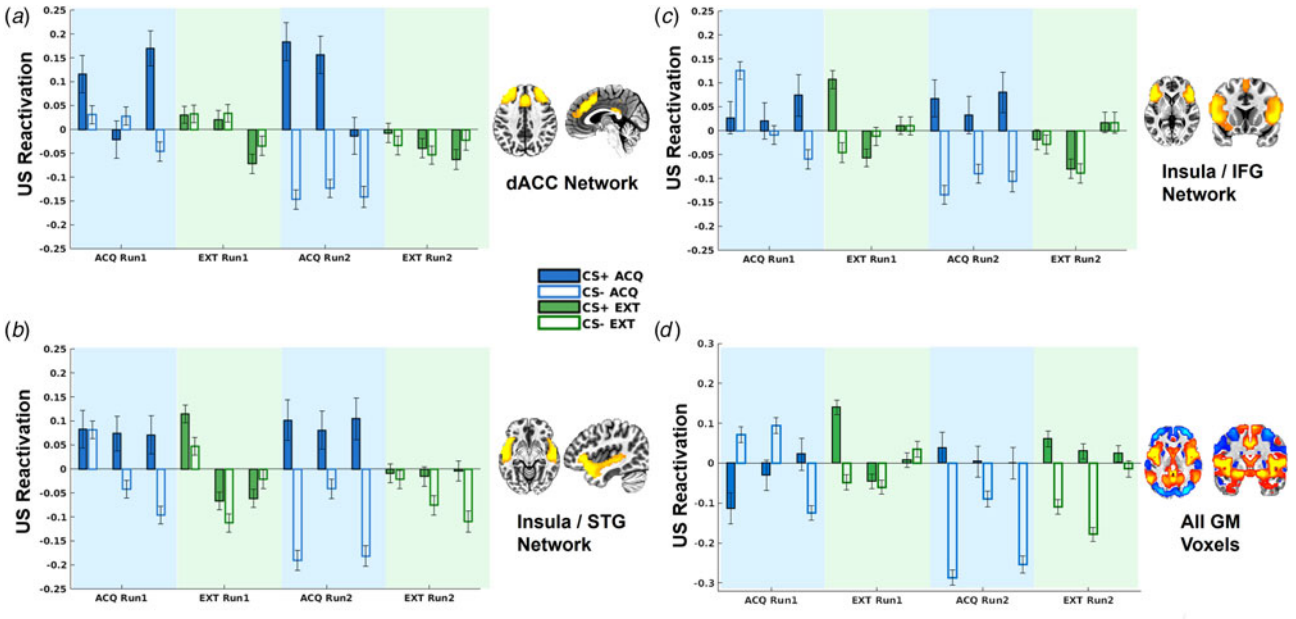

Figure 3. US decoder output (i.e. hyperplane distances) for the dorsal ACC network (a), insula/STG network (b), insula/IFG network (c), and whole-brain grey matter (GM) mask (d). Data are separated into early, middle, and late blocks within each phase.

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

interactive term for block in the LMEMs. This analysis demonstrated that US decoder output for the dACC and insula/STG networks was not moderated by block (*ps* > 0.15). By contrast, the insula/MFG network demonstrated a significant CS × context × block interaction, $t(8873) = 2.79$, $p = 0.0053$, such that greater US decoder output for the CS+ *v.* CS− in the acquisition *v.* extinction context was only observed in the last half of blocks, $t(4444) = 2.71$, $p = 0.0067$, and not first half of blocks, $t(4424) = -0.44$, $p = 0.66$ (Fig. 3). Similarly, the whole-brain GM mask demonstrated a significant CS × context × block interaction, $t(8873) = 4.17$, $p < 0.001$, such that greater US decoder output for the CS+ *v.* CS− in the acquisition *v.* extinction context was observed in the last half of blocks, $t(4444) = 2.05$, $p = 0.04$, whereas this pattern was inverted in the first half of blocks, $t(4424) = -2.25$, $p = 0.025$ (Fig. 3).

CS reactivation of US representations and convergent fear responses during day 1

Trial-by-trial SCRs demonstrated the predicted CS × context interaction, $t(8505) = 9.32$, $p < 0.001$ (online Supplementary Fig. S3). US decoder output for the insula/STG network was significantly positively correlated with trial-by-trial SCRs, $t(8262) = 3.06$, $p = 0.002$, and no interaction with CS or context (Fig. 4). US decoder output for the whole-brain GM mask was similarly positively correlated with trial-by-trial SCRs, $t(8262) = 7.55$, $p < 0.001$ (Fig. 4).

US expectancy ratings were also available for day 1. US decoder output for the insula/STG network was marginally significantly related to the context × rating interaction, $t(8873) = 2.59$, $p = 0.0096$, such that US decoder output was strongly related to expectancy ratings during acquisition, $t(3791) = 3.49$, $p < 0.001$, but not extinction contexts, $t(5077) = -1.13$, $p = 0.26$ (online Supplementary Fig. S5). US decoder output for the other networks was not related to expectancy ratings during day 1.

CS reactivation of US representations during day 2 initial fear recall and impact of L-DOPA

On day 2, there was a significant L-DOPA *v.* placebo × context interaction for US decoder output for the insula/STG network, $t(581) = 3.17$, $p = 0.002$ (Fig. 5). This interaction was attributed to greater US reactivation in the extinction context among the placebo group relative to the L-DOPA group $t(287) = -2.42$, $p = 0.016$. Further, there was greater US reactivation in the extinction relative to acquisition context in the placebo group, $t(225) = -2.87$, $p = 0.004$, which was absent in the L-DOPA group, $t(351) = 1.33$, $p = 0.18$.

CS reactivation of US representations during day 2 initial fear recall, skin conductance responses, and impact of L-DOPA

There was no impact of L-DOPA on SCR responses during day 2 initial fear recall (online Supplementary Fig. S4). Mirroring the correlation observed on day 1, trial-by-trial SCRs were predictive of US decoder output for the insula/STG network, $t(499) = 3.32$, $p < 0.001$ (Fig. 4), with no differences between L-DOPA groups.

CS reactivation of US representations during day 2 fear reinstatement and impact of L-DOPA

There was no impact of L-DOPA on US decoder output following fear reinstatement for any network.

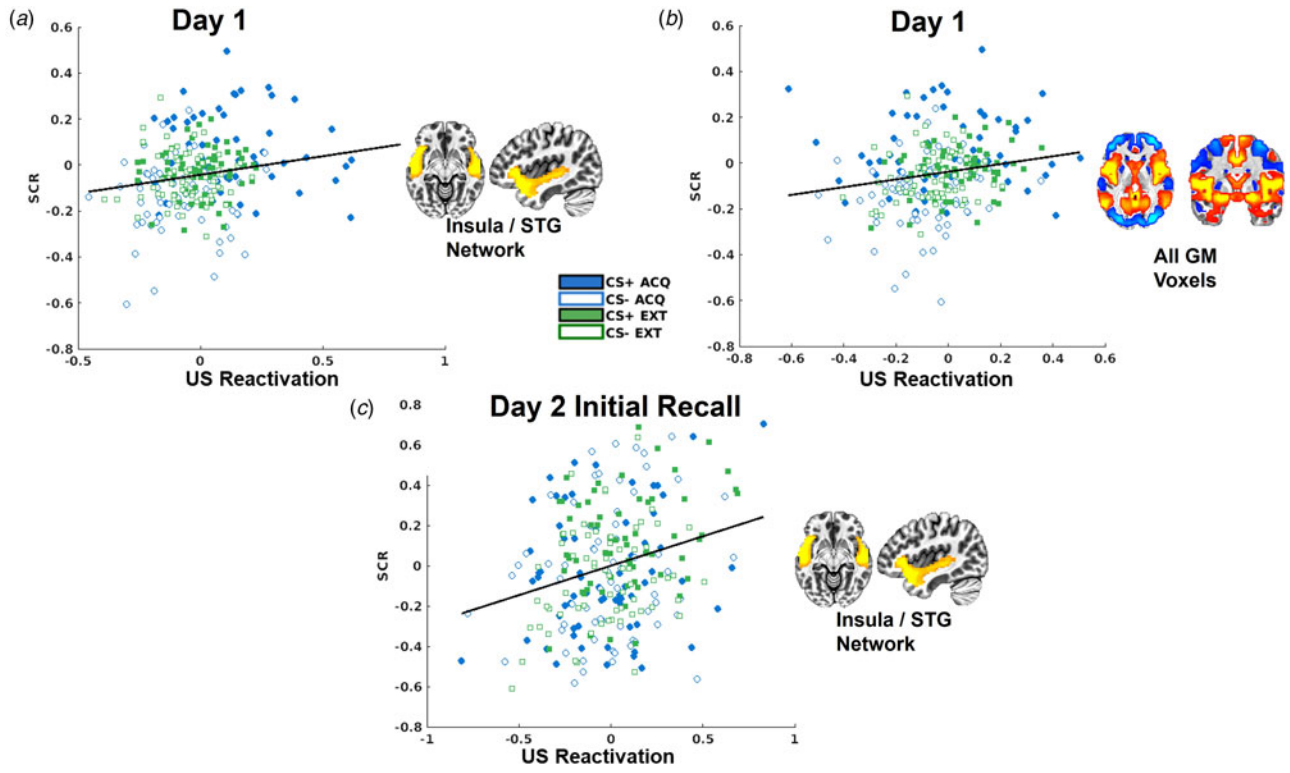

Figure 4. Scatterplots depicting the relationships between skin conductance responses (SCRs) and US decoder output (i.e. hyperplane distances) for the insula/STG network for day 1 (a), for the whole-brain GM mask (b), and for the insula/STG network during the initial fear recall test on day 2 (c).

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

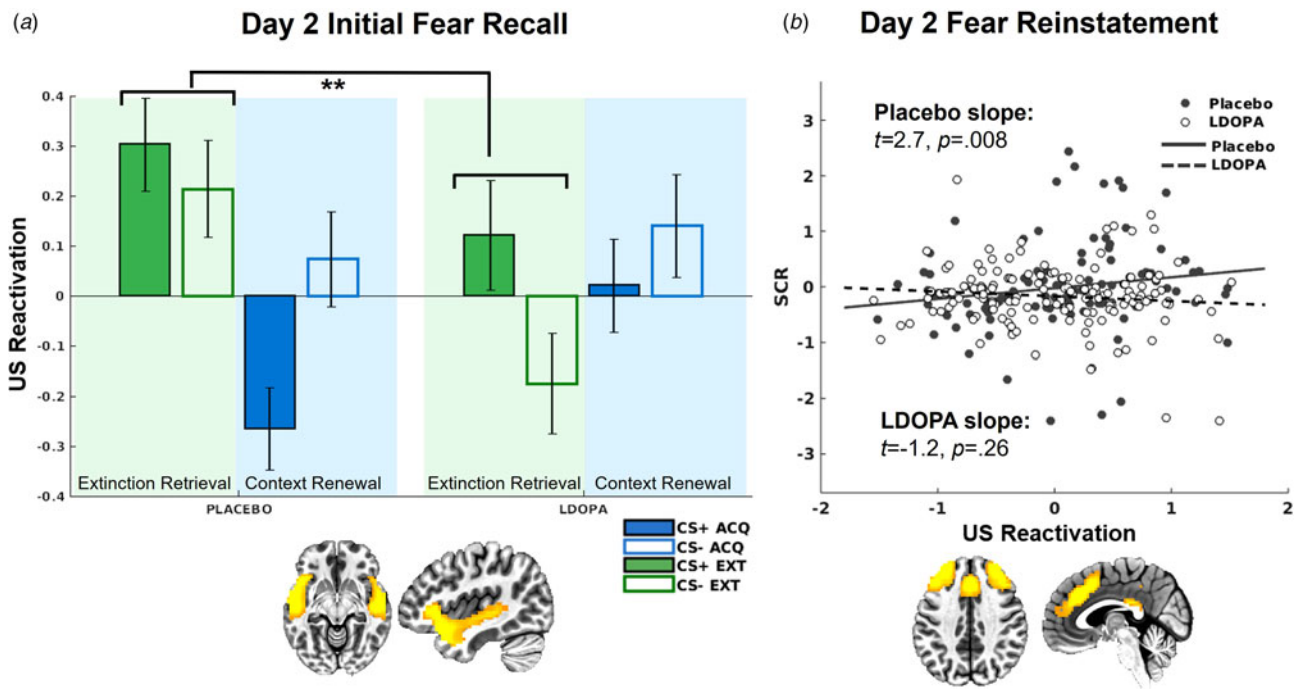

Figure 5. (a) US decoder output (i.e. hyperplane distances) for the insula/STG network, indicating significantly greater US representations in the extinction context among the placebo group compared to the L-DOPA group. (b) US decoder output (i.e. hyperplane distances) for the dACC network following reinstatement only positively predicted trial-by-trial SCRs for the placebo group, but not L-DOPA group.

CS reactivation of US representations during day 2 fear reinstatement, skin conductance responses, and impact of L-DOPA

L-DOPA decreased SCRs following fear reinstatement compared to placebo, $t(3184) = -2.23$, $p = 0.026$ (online Supplementary Fig. S4). US decoder output for the insula/STG network was again significantly related to trial-by-trial SCRs following reinstatement, $t(986) = 2.79$, $p = 0.005$, with no significant differences between groups controlling for multiple comparisons, $t(986) = -2.26$, $p = 0.024$. There was a reinstatement × SCR × L-DOPA *v.* placebo interaction for the dACC network, $t(986) = -3.17$, $p = 0.002$ (Fig. 5), such that following reinstatement, SCRs were only predictive of dACC US decoder output among the placebo group, $t(203) = 2.68$, $p = 0.008$, but not the L-DOPA group, $t(274) = -1.12$, $p = 0.26$. SCRs were not predictive of US decoder output following reinstatement in any other network.

Discussion

The purpose of the present study was to translate a novel measure of fear memory strength recently demonstrated in animal models to a human neuroimaging context as a test of the impact of dopaminergic manipulations on extinction consolidation among women with PTSD. With respect to translating the hypothesis that, during fear learning, the CS reactivates neural representations of the US (Grewe et al., 2017; Herry & Jercog, 2022; Luo et al., 2018), results among these neuroimaging data were consistent with predictions. We used MVPA to define a US representation, then tested whether the CS+ reactivates this US representation during fear learning and ceases to reactivate the US representation during fear extinction. We found support for the predicted US reactivations on day 1 in networks consisting of dorsal anterior cingulate and lateral PFC, and insular cortex and superior temporal gyrus. Meta-analyses of pain processing generally corroborate the importance of these regions for processing painful US (Biggs et al., 2020; Palermo et al., 2015; Xu et al., 2020). While these meta-analyses further demonstrate that pain processing is widely spatially distributed across the brain, and indeed classifier accuracy was above chance for all networks tested here and substantially highest for the whole-brain GM mask, the CS only reactivated US representations in these two networks. This suggests a dissociation in the brain networks processing the delivery of the US and the networks in which the CS is able to elicit reactivations of US representations during associative learning. That is, the CS does not appear to reactivate the brain-wide US representation; rather, the CS appears to reactive the US representation selectively within specific networks.

An additional network consisting of anterior insula and inferior frontal and middle frontal gyri, as well as the whole-brain GM mask, demonstrated delayed reactivation of US representations, with US representations more likely to reactive in the predicted pattern during later blocks of the task. Further investigation of the temporal dynamics in which US reactivation occurs at various points during fear acquisition and extinction across differing networks is warranted. Suggesting validity of the US representations in response to the CS in the insula/STG network, US decoder output was positively correlated with trial-by-trial SCRs on day 1 and day 2 as well as with expectancy ratings in the acquisition context on day 1 (online Supplementary Fig. S5). The lack of amygdala involvement in this study echoes the broader human fear conditioning and extinction literature (Fullana et al., 2016, 2018; Morriss, Hoare, & van Reekum, 2018). Alternatively, the lack of amygdala involvement observed here could be due to subtle timing issues or specific subnuclei of the amygdala that are difficult to resolve

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

with the spatial and temporal resolution generally available in standard fMRI methodology (Wen et al., 2022).

We observed that women who received L-DOPA following day 1 context learning demonstrated decreased US reactivation in the insula/STG network relative woman who received placebo uniquely in the extinction context (i.e. during the test of extinction retrieval). No differences in US reactivation were observed between groups following initial context renewal (i.e. return to the fear acquisition context); however, only the placebo group demonstrated significantly greater US reactivation in the extinction relative to acquisition context. Suggesting validity of US decoder output on day 2, this MVPA measure of CS reactivation of US representations was again correlated with trial-by-trial SCRs, further supporting the interpretation of heightened fear memory strength in the extinction context among the placebo group and lack thereof in the L-DOPA group.

On the one hand, the temporal ordering of the experiment was such that the extinction retrieval test occurred first, followed by the context renewal test. As such, the L-DOPA group may have demonstrated increased extinction memory strength relative to fear memory strength in the extinction context, as indexed by US reactivation, due to the hypothesized role of increased dopamine signaling during consolidation (Gerlicher et al., 2018; Haaker et al., 2013; Kalisch et al., 2019; Luo et al., 2018). The lack of differences following context renewal could simply be due to the fact that context renewal occurred following the extinction retrieval test, which essentially functioned as another extinction training block and led to less fear memory strength relative to extinction memory strength in the acquisition context in both groups.

On the other hand, the pattern of findings could be related to dysfunctional context processing in PTSD and suggest that L-DOPA and the associated increased dopamine signaling during consolidation of context-dependent fear and extinction memories rescues this dysfunctional context processing. Recent models posit altered processing of context as a core mechanism underlying PTSD (Liberzon & Abelson, 2016). As an example, a prior study among adults with PTSD using a contextual fear conditioning and extinction paradigm demonstrated greater fear reinstatement 24 h later in the fear extinction context compared to the fear acquisition context (Garfinkel et al., 2014). Possibly consistent with this hypothesis of altered context processing in PTSD, results demonstrated greater US decoder output for the insula/STG network in the extinction context relative to the acquisition context among women who received the placebo, an effect which was absent in the L-DOPA groups. Additional research is necessary to disentangle the unique impacts of L-DOPA on extinction retrieval (Gerlicher et al., 2018) *v.* context renewal (Haaker et al., 2013) *v.* altered context processing in PTSD (Liberzon & Abelson, 2016).

Following fear reinstatement, while we observed decreased SCRs, relative to before reinstatement, among the L-DOPA group (as we previously reported, Cisler et al. [2020]), we did not observe differences in US decoder output for any network between groups. However, we did observe altered relationships with SCRs between groups, such that US decoder output in the dACC and lateral PFC network was only related to trial-by-trial SCRs in the placebo group, and not L-DOPA group. This might suggest that US decoder output in the placebo group in this network was tracking fear memory strength following the fear reinstatement procedure, consistent with an increase in fear memory strength following reinstatement. By contrast, in the L-DOPA groups, a lack of reinstated fear memory did not result in the absence of US decoder output; rather, the US decoder output may have no longer been tracking fear memory strength. That is, trial-by-trial changes in voxel activity within this network would necessarily result in trial-by-trial changes in US decoder output; however, those trial-by-trial US decoder outputs do not necessarily have to reflect fear memory strength. One possible explanation is that there is a floor effect, such that below some degree of true fear memory strength, and hence valid US decoder output, the trial-by-trial changes in decoder output reflect noise or other sources of variance in the network's voxelwise activity patterns. Here, the placebo group's fear memory strength may have been low (e.g. due to the repeated extinction on day 2), but sufficient to still produce US decoder output that meaningfully tracks fear memory strength, whereas the L-DOPA group's fear memory strength may have been below a requisite magnitude to produce US decoder output that meaningfully tracks fear memory strength. This interpretation is speculative and further testing of this novel MVPA method for tracking fear memory strength is needed, particularly with testing convergent relationships with other fear memory strength measures (e.g. pupillometry).

Data from this study support the hypothesis that a fear CS reactivates neural representations of an aversive US and provides additional evidence for the emerging literature suggesting that increased dopaminergic signaling during consolidation enhances fear extinction memory strength, thereby leading to decreased fear memory strength (Bouchet et al., 2017; Gerlicher et al., 2018; Haaker et al., 2013; Kalisch et al., 2019; Luo et al., 2018). Nonetheless, this study is not without limitations. First, this is a reanalysis of our prior RCT data and therefore should be considered a post-hoc analysis; interpretations regarding the impact of L-DOPA on inhibiting fear memory retrieval in the extinction context should be tempered accordingly. Second, the sample was limited to adult women with PTSD, and comparisons of CS reactivation of US neural representations in healthy samples are needed. For example, it is plausible, given differences in PTSD in fear acquisition and extinction learning (Suarez-Jimenez et al., 2020), that different networks represent the US in PTSD *v.* healthy samples. Third, we adopted the context conditioning task used in a prior L-DOPA trial (Haaker et al., 2013) that conducted both fear acquisition and extinction on day 1, then fear and extinction retrieval on day 2. Further testing of US decoder output is needed using 3-day designs to isolate better the fear acquisition, extinction, and retrieval phases. Fourth, as noted, further testing of US decoder output as a measure of fear memory strength is needed with additional measures of fear responding (Ojala & Bach, 2020). Fifth, the aversive US here was a painful electric shock, and it remains to be seen the degree to which US reactivation occurs for other types of aversive US. Finally, while our results support US reactivation as a measure of fear memory strength, it is relevant to note that this measure is agnostic with respect to understanding or quantifying whether an extinction procedure erases/decays the fear memory *v.* the formation of the extinction memory inhibits retrieval of the fear memory.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291723002891.

Funding statement. J. M. C. is supported by MH119132, MH108753, and AA030740. J. E. D. is supported by MH122387. G. A. J. is supported by DA048022.

Competing interests. None.

https://doi.org/10.1017/S0033291723002891 Published online by Cambridge University Press

References

Bach, D. R., & Melinscak, F. (2020). Psychophysiological modelling and the measurement of fear conditioning. *Behaviour Research and Therapy*, *127*, 103576. doi: 10.1016/j.brat.2020.103576

Biggs, E. E., Timmers, I., Meulders, A., Vlaeyen, J. W. S., Goebel, R., & Kaas, A. L. (2020). The neural correlates of pain-related fear: A meta-analysis comparing fear conditioning studies using painful and non-painful stimuli. *Neuroscience & Biobehavioral Reviews*, *119*, 52–65. doi: 10.1016/j.neubiorev.2020.09.016

Bouchet, C. A., Miner, M. A., Loetz, E. C., Rosberg, A. J., Hake, H. S., Farmer, C. E., … Greenwood, B. N. (2017). Activation of nigrostriatal dopamine neurons during fear extinction prevents the renewal of fear. *Neuropsychopharmacology*, *43*(3), 665–672. doi: 10.1038/npp.2017.235

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, *11*(5), 485–494. doi: 10.1101/lm.78804

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, *14*(3), 140–151.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1–27:27. doi: 10.1145/1961189.1961199

Cisler, J. M., Privratsky, A. A., Sartin-Tarm, A., Sellnow, K., Ross, M., Weaver, S., … Kilts, C. D. (2020). L-DOPA and consolidation of fear extinction learning among women with posttraumatic stress disorder. *Translational Psychiatry*, *10*(1), 1–11.

Cisler, J. M., Tamman, A. J. F., & Fonzo, G. A. (2023). Diminished prospective mental representations of reward mediate reward learning strategies among youth with internalizing symptoms. *Psychological Medicine*, 1–11. doi: 10.1017/S0033291723000478

Crombie, K. M., Azar, A., Botsford, C., Heilicher, M., Moughrabi, N., Gruichich, T. S., … Cisler, J. M. (2023). Aerobic exercise after extinction learning reduces return of fear and enhances memory of items encoded during extinction learning. *Mental Health and Physical Activity*, *24*, 100510. doi: 10.1016/j.mhpa.2023.100510

Crombie, K. M., Sartin-Tarm, A., Sellnow, K., Ahrenholtz, R., Lee, S., Matalamaki, M., … Cisler, J. M. (2021). Aerobic exercise and consolidation of fear extinction learning among women with posttraumatic stress disorder. *Behaviour Research and Therapy*, *142*, 103867.

Dunsmoor, J. E., Cisler, J. M., Fonzo, G. A., Creech, S. K., & Nemeroff, C. B. (2022). Laboratory models of post-traumatic stress disorder: The elusive bridge to translation. *Neuron*, *110*(11), 1754–1776.

Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology*, *59*(5), 715–723.

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., … Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience and Biobehavioral Reviews*, *88*, 16–25. doi: 10.1016/j.neubiorev.2018.03.002

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508.

Garfinkel, S. N., Abelson, J. L., King, A. P., Sripada, R. K., Wang, X., Gaines, L. M., & Liberzon, I. (2014). Impaired contextual modulation of memories in PTSD: An fMRI and psychophysiological study of extinction retention and fear renewal. *The Journal of Neuroscience*, *34*(40), 13435–13443. doi: 10.1523/JNEUROSCI.4287-13.2014

Gerlicher, A. M. V., Tüscher, O., & Kalisch, R. (2018). Dopamine-dependent prefrontal reactivations explain long-term benefit of fear extinction. *Nature Communications*, *9*(1), 1–9. doi: 10.1038/s41467-018-06785-y

Grewe, B. F., Gründemann, J., Kitch, L. J., Lecoq, J. A., Parker, J. G., Marshall, J. D., … Schnitzer, M. J. (2017). Neural ensemble dynamics underlying a long-term associative memory. *Nature*, *543*(7647), 670–675. doi: 10.1038/nature21682

Haaker, J., Gaburro, S., Sah, A., Gartmann, N., Lonsdorf, T. B., Meier, K., … Kalisch, R. (2013). Single dose of l-dopa makes extinction memories context-independent and prevents the return of fear. *Proceedings of the National Academy of Sciences*, *110*(26), E2428–E2436. doi: 10.1073/pnas.1303061110

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Hennings, A. C., Cooper, S. E., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2022). Pattern analysis of neuroimaging data reveals novel insights on threat learning and extinction in humans. *Neuroscience and Biobehavioral Reviews*, *142*, 104918. doi: 10.1016/j.neubiorev.2022.104918

Herry, C., & Jercog, D. (2022). Decoding defensive systems. *Current Opinion in Neurobiology*, *76*, 102600. doi: 10.1016/j.conb.2022.102600

Kalisch, R., Gerlicher, A. M. V., & Duvarci, S. (2019). A dopaminergic basis for fear extinction. *Trends in Cognitive Sciences*, *23*(4), 274–277. doi: 10.1016/j.tics.2019.01.013

Knudsen, E. I. (1994). Supervised learning in the brain. *Journal of Neuroscience*, *14*(7), 3985–3997. doi: 10.1523/JNEUROSCI.14-07-03985.1994

Liberzon, I., & Abelson, J. L. (2016). Context processing and the neurobiology of post-traumatic stress disorder. *Neuron*, *92*(1), 14–30. doi: 10.1016/j.neuron.2016.09.039

Luo, R., Uematsu, A., Weitemier, A., Aquili, L., Koivumaa, J., McHugh, T. J., & Johansen, J. P. (2018). A dopaminergic switch for fear to safety transitions. *Nature Communications*, *9*(1), 2483. doi: 10.1038/s41467-018-04784-7

Morriss, J., Hoare, S., & van Reekum, C. M. (2018). It's time: A commentary on fear extinction in the human brain using fMRI. *Neuroscience & Biobehavioral Reviews*, *94*, 321–322. doi: 10.1016/j.neubiorev.2018.06.025

Moughrabi, N., Botsford, C., Gruichich, T. S., Azar, A., Heilicher, M., Hiser, J., … Cisler, J. M. (2022). Large-scale neural network computations and multivariate representations during approach-avoidance conflict decision-making. *NeuroImage*, *264*, 119709. doi: 10.1016/j.neuroimage.2022.119709

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. doi: 10.1016/j.neuroimage.2011.08.076

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neuroscience & Biobehavioral Reviews*, *114*, 96–112. doi: 10.1016/j.neubiorev.2020.04.019

Palermo, S., Benedetti, F., Costa, T., & Amanzio, M. (2015). Pain anticipation: An activation likelihood estimation meta-analysis of brain imaging studies. *Human Brain Mapping*, *36*(5), 1648–1661. doi: 10.1002/hbm.22727

Raij, T., Nummenmaa, A., Marin, M.-F., Porter, D., Furtak, S., Setsompop, K., & Milad, M. R. (2018). Prefrontal cortex stimulation enhances fear extinction memory in humans. *Biological Psychiatry*, *84*(2), 129–137. doi: 10.1016/j.biopsych.2017.10.022

Rothbaum, B. O., & Davis, M. (2003). Applying learning principles to the treatment of post-trauma reactions. *Annals of the New York Academy of Sciences*, *1008*, 112–121.

Suarez-Jimenez, B., Albajes-Eizagirre, A., Lazarov, A., Zhu, X., Harrison, B. J., Radua, J., … Fullana, M. A. (2020). Neural signatures of conditioning, extinction learning, and extinction recall in posttraumatic stress disorder: A meta-analysis of functional magnetic resonance imaging studies. *Psychological Medicine*, *50*(9), 1442–1451. doi: 10.1017/S0033291719001387

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-Administered PTSD Scale: A review of the first ten years of research. *Depression and Anxiety*, *13*(3), 132–156.

Wen, Z., Raio, C. M., Pace-Schott, E. F., Lazar, S. W., LeDoux, J. E., Phelps, E. A., & Milad, M. R. (2022). Temporally and anatomically specific contributions of the human amygdala to threat and safety learning. *Proceedings of the National Academy of Sciences*, *119*(26), e2204066119. doi: 10.1073/pnas.2204066119

Xu, A., Larsen, B., Baller, E. B., Scott, J. C., Sharma, V., Adebimpe, A., … Satterthwaite, T. D. (2020). Convergent neural representations of experimentally-induced acute pain in healthy volunteers: A large-scale fMRI

meta-analysis. *Neuroscience & Biobehavioral Reviews*, *112*, 300–323. doi: 10.1016/j.neubiorev.2020.01.004

Zaki, Y., Mau, W., Cincotta, C., Monasterio, A., Odom, E., Doucette, E., … Ramirez, S. (2022). Hippocampus and amygdala fear memory engrams re-emerge after contextual fear relapse. *Neuropsychopharmacology*, *47*(11), 1992–2001. doi: 10.1038/s41386-022-01407-0

Zhang, X., & Li, B. (2018). Population coding of valence in the basolateral amygdala. *Nature Communications*, *9*(1), 5195. doi: 10.1038/s41467-018-07679-9

Zhou, F., Zhao, W., Qi, Z., Geng, Y., Yao, S., Kendrick, K. M., … Becker, B. (2021). A distributed fMRI-based signature for the subjective experience of fear. *Nature Communications*, *12*(1), 6643. doi: 10.1038/s41467-021-26977-3