**fMRI Tasks**

**Fear Conditioning, Fear Extinction, and Fear Recall Task**

The task used here was modeled after a prior study among healthy adults testing the impact of L-DOPA on fear context renewal(1). The unconditioned stimulus (US) used during Day 1 was an electric shock. Participants calibrated US intensity to a level that was uncomfortable but not painful (i.e., a rating of 7 on a 10 point Likert scale). Day 1 US stimulator output (M = .24, SD = .07) did not differ between groups, all *p*s > 0.375. Conditioned stimuli consisted of triangles and circles, each displayed for 3s with a jittered inter-trial interval of 2-6s. Colored backgrounds distinguished the acquisition and extinction contexts. The stimuli serving as CS+ vs CS- and colors distinguishing contexts were counterbalanced across participants. An initial baseline phase consisted of 6 presentations of each stimulus with no UCS onsets. The task then alternated between acquisition and extinction phases, with two presentations of each phase. The acquisition phase presented each CS 18 times, with a shock occurring 2.5s following CS+ presentation with a 50% reinforcement schedule. The extinction phase presented each stimulus 18 times and no shocks occurred. There were a total of 156 trials. Participants’ instruction on the task was to indicate the identity of the stimulus and were not informed about any specific contingencies between stimuli and shocks. US expectancy assessment occurred after every 12 trials (i.e., three contingency awareness assessments per context), in which participants provided a 0-10 rating of how likely they believed the shock was to follow each of the stimuli.

Participants returned for a recall test 24 hrs later using a similar procedure and task design. Consistent with the prior L-DOPA study among healthy participants(1), participants first recalibrated the US intensity again to a 7 (“uncomfortable but not painful”) on a 10 point Likert scale. Day 2 US stimulator output did not differ between groups, all *p*s > 0.355. Consistent with the prior study, the recall task alternated between acquisition and extinction contexts, in a pseudorandom order for a total of 6 context repetitions, and presented two CS+ and CS- stimuli per context presentation (i.e., 12 total CS+ and CS- repetitions). There were no shocks delivered during the task during Day 2. Each stimulus was again presented for 3s with a jittered inter-trial interval of 2-6s. After the first run of the task, participants then received a single uncued US presentation (i.e., reinstatement), and then participants completed again the identical task as implemented in the first run. Within both runs, additional generalization stimuli were presented that were not used during the Day 1 learning task, for which subsequent manuscripts will focus on describing and presenting results. US expectancy was collected on Day 2, but unfortunately the timing of the assessments was not done in a way that allowed differentiation of contexts and therefore the data are uninterpretable.

**Linear Mixed Effects Models for US expectancy on Day 1.** Parallel LMEMs were conducted on the US expectancy ratings as those reported for SCR in the main manuscript. Here, because the US expectancy ratings occurred every 12 trials (i.e., 3 times per block), we modeled the expectancy rating across the trial-by-trial US decoder output for the preceding trials (e.g., a US expectancy rating of 5/10 provided after trial 12 was modeled across the 12 preceding trials). Accordingly, this analysis tests if US decoder output is higher (or lower) on blocks of trials were the individual had higher (or lower) US expectations.

**Double Blind Randomization**

Participants were randomized using blocked stratified randomization, in which randomization was stratified based on age (>35 or <= 35), number of comorbid diagnoses (<3 or >=3), and whether they were currently prescribed psychotropic medication (yes or no). Randomization was successful in creating groups balanced on key demographic and clinical variables (see Table 1 and supplemental Table S1).

**Independent Component Analysis**

An Independent Component Analysis (Calhoun *et al.*, 2001) (ICA) with a model order of 35 components was conducted on the full voxelwise fMRI timecourses. This model order delivered a good balance between component reliability estimated across 50 ICASSO iterations and interpretability of canonical networks. 13 of the 35 networks were considered functional networks after removing noise networks (e.g., CSF, WM, head motion) and networks of non-interests (e.g., primary motor networks, visual networks). 5 of these 13 components were selected a priori for analysis in this study based on prior meta-analyses of fear conditioning and pain processing (Palermo *et al.*, 2015; Fullana *et al.*, 2016, 2018; Biggs *et al.*, 2020; 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 grey matter (GM) voxels to define a whole-brain GM US reactivation model.

**Sample Size**

Sample was determined via power analysis based on effect sizes of SCR data available from a prior study among healthy humans(1).

**Skin conductance acquisition and processing**

At both sites, SCR data were acquired on a BIOPAC MP150 Data Acquisition System using the EDA100C module with MECMRI-TRANS cable system. Data were acquired directly into BIOPAC AcqKnowledge 4.3 software at 2000 Hz (Arkansas site) or 1000 Hz (Wisconsin site). Shocks were administered via the BIOPAC STM100C module using pre-gelled electrodes placed on the skin of the fleshy portion of the mediolateral, left lower leg, directly over the tibialis anterior. SCR recording electrodes were placed on the medial portions of the thenar and hypothenar eminences of the left hand; ground electrode was placed on the ventral surface of the left wrist. Amperage on the stimulation device was set to the maximum (50 mA) to allow the greatest range of intensity selections. Participants were told to select an intensity of a 7/10 pain scale.

Consistent with prior studies(1,2), prior to breaking of the blind and performing any analyses, participants whose Day 2 SCR data showed excessive artifact or flat responding were removed from Day 2 SCR analyses (total n=19; n=8 from placebo, n=6 from 100mg, n=5 from 200mg). This amount of data loss (22%) is commensurate with prior fear extinction studies using SCR(1–3). Skin conductance data underwent preprocessing consistent with contemporary recommendations (4–8,8–10), which included, in order, 1) a 10ms median filter, 2) unidirectional butterworth filter with .0159hz and 5hz low and high pass frequencies, and 3) downsampling to 10hz. Skin conductance responses were then estimated on a trial-by-trial basis by applying the well-validated forward convolution model of skin conductance responses within a GLM approach(6,8–11). Resulting SCRs were normalized to each individual’s max SCR per day to account for inter-individual differences in overall magnitude of SCR responding. Reinforced CS+ trials from Day 1 Acquisition phase blocks were not included in analyses to avoid any contamination of SCR responses to the stimulus with SCR responses to the shock.

**MRI conductance acquisition and processing**

At the Arkansas site, fMRI data were acquired on a Philips Achieva 3T X-series scanner using a 32-channel headcoil. T1-weighted anatomic images were acquired with a MP-RAGE sequence (matrix = 192 × 192, 160 sagittal slices, TR/TE/FA = 7.5/3.7/9°, FOV = 256, 256, 160, final resolution = 1 × 1 × 1 mm resolution). Echo planar imaging sequences were used to collect the functional images using the following sequence parameters: TR/TE/FA = 2000 ms/30 ms/90°, FOV = 240 × 240 mm, matrix = 80 × 80, 37 axial slices (parallel to AC–PC plane to minimize OFC signal artifact), slice thickness = 2.5mm, and final resolution of 3 × 3 × 3 mm.

At the UW-Madison site, fMRI data were acquired on a GE MR750 3T scanner using an 8-channel headcoil. T1-weighted anatomic images were acquired with a MP-RAGE sequence (matrix = 256x256, 156 axial slices, TR/TE/FA = 8.2ms/3.2ms/12°, FOV = 25.6cm, final resolution = 1x1x1mm). EPI sequences used to collect the functional images used the following parameters: TR/TE/FA = 2000ms/ 25 ms/ 60, FOV = 24cm, matrix = 64 x 64, 40 sagittal slices, slice thickness = 4mm, original resolution was 4 x 3.75 x 3.75, and images were resampled to match the resolution of the UAMS data of 3x3x3mm.

Image preprocessing followed standard steps and was completed using AFNI software. In the following order, images underwent despiking, slice timing correction, deobliquing, motion correction using rigid body alignment, alignment to participant’s normalized anatomical images, spatial smoothing using a 8 mm FWHM Gaussian filter (AFNIs 3dBlurToFWHM that estimates the amount of smoothing to add to each dataset to result in the desired level of final smoothing), detrending, bandpass filtering (low frequency [.0078 Hz] for task data and low and high frequency [.01Hz - .1Hz] for resting-state data), and rescaling into percent signal change. Images were normalized using the MNI 452 template brain. We corrected for head motion related signal artifacts by using motion regressors derived from Volterra expansion, consisting of [R R2 Rt-1 R2t-1], where R refers to each of the 6 motion parameters, and separate regressors for mean signal in the CSF and WM. This step was implemented directly after motion correction and normalization of the EPI images in the image preprocessing stream. Additionally, we censored TRs from the first-level GLMs based on threshold of framewise displacement (FD) > 0.4. FD refers to the sum of the absolute value of temporal differences across the 6 motion parameters; thus, a cut-off of 0.4 results in censoring TRs where the participant moved, in total across the 6 parameters, more than ~0.4 mm plus the immediately following TR (to account for delayed effects of motion artifact). Additionally, we censored isolated TRs where the preceding and following TRs were censored, and we censored entire runs if more than 50% of TRs within that run were censored. This led to the removal of 10 women from task analyses (n=4 from placebo, n=3 from 100mg, n=3 from 200mg) and 7 women (n=1 from placebo, n=1 from 100mg, n=5 from 200mg) from resting-state analyses due to missing data.

**Correcting for Site Differences in Imaging Data**

Differences between sites in trial-by-trial voxelwise beta coefficients (from the trial-by-trial least squares models (AFNIs 3dLSS) ) were corrected using combat (12,13). To ensure separation of data between model building and model testing, we conducted three sets of site corrections using combat. First, the 3dLSS trial-by-trial voxelwise betas for the US vs the no-US data were corrected for site. Second, the 3dLSS trial-by-trial voxelwise betas for each CS during Day 1 were corrected for site. Third, the 3dLSS trial-by-trial voxelwise betas for each CS during Day 2 were corrected for site.

**Randomization Implementation**

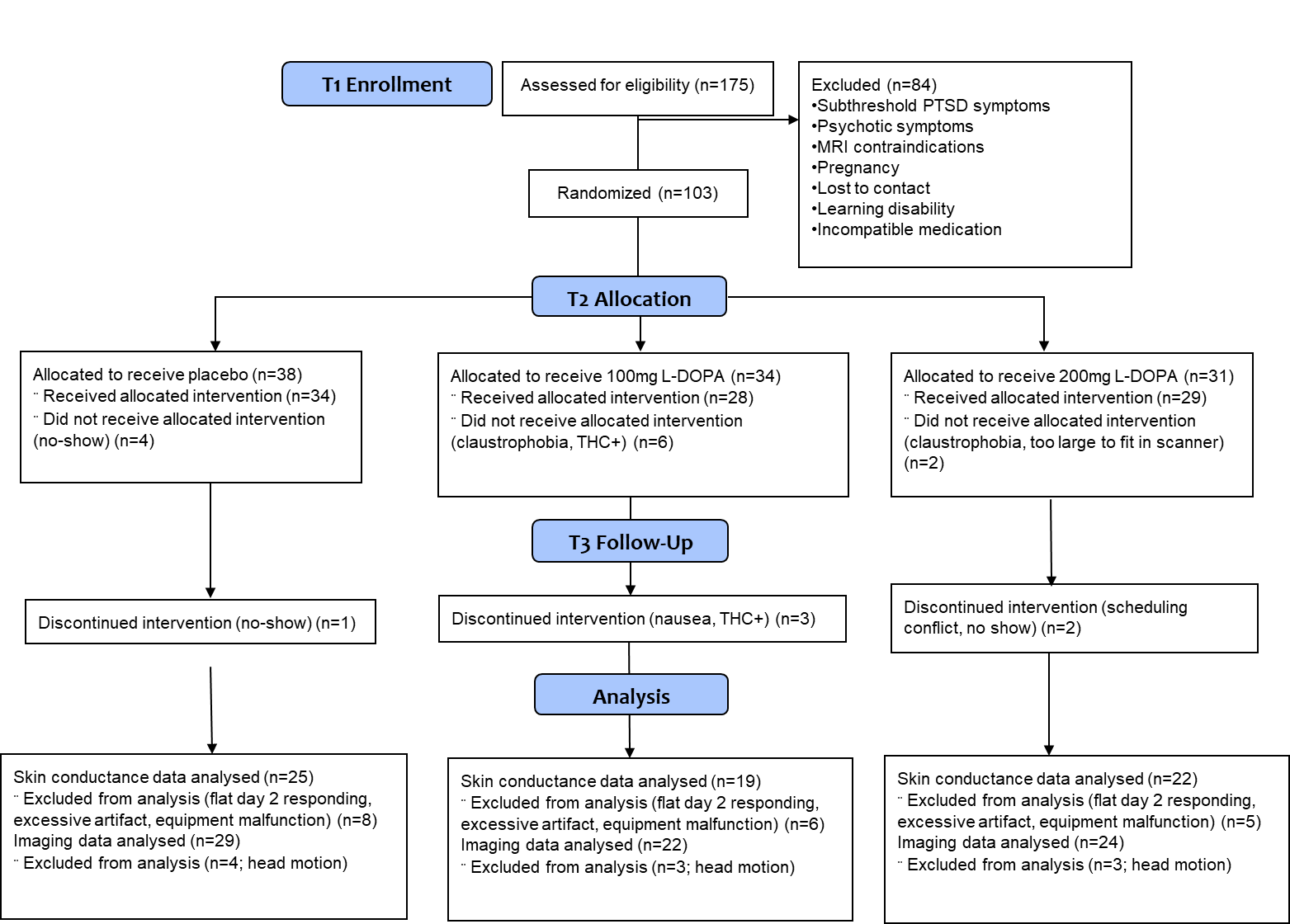
Blocked and stratified randomization sequences were generated using custom code in Matlab. The randomization sequence and blind was maintained by an independent pharmaceutical research center. Placebo and L-DOPA were in capsule form and identical in appearance. Pills were prepacked in bottles by the independent pharmaceutical research center, and consecutively numbered for each woman according to the randomization schedule. Both participants and research staff were blind to drug allocation. Study enrolment occurred between April 2016 and May 2018 and ended after the targeted sample size was achieved. The study protocol can accessed by contacting the corresponding author.

Supplemental Table 1. Birth control, estrogen levels, smoking status, and psychiatric medication usage across the drug groups among all randomized participants.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Placebo (a)**  **N=34** | **100mg (b)**  **N=28** | **200mg (c)**  **N=29** | ***p* values** |
| **Birth control (%)** | 50.0 | 53.6 | 48.3 | p(abc) = 0.920 |
| **Estradiol concentration\* (pg/mL)** | 1.45 (.82) | 1.43 (.71) | 1.34 (.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 |
| **Receiving psychotherapy (%)** | 20.9 | 15.4 | 16.5 | p(abc) = 0.891 |
| **Psychotropic medication (%)**    **Antidepressants (%)**    **SSRI (%)**    **SNRI (%)**    **NDRI (%)**  **Mood**  **stabilizer/antipsychotic (%)**    **Benzo (%)**    **Stimulants (%)**  **Anticonvulsant (%)**  **Antianxiety (%)**  **DA receptor antagonist (%)**  **DA receptor agonist (%)**    **Any psychotropic med (%)** | 17.6    26.5    2.9    11.8    5.9    14.7    11.8    17.6  2.9  2.9  8.8  58.8 | 25.0    17.9    10.7    14.3    10.7  3.6    3.6  10.7  0.0  7.1  7.1  50.0 | 27.6    27.6    10.3  10.3  13.8  10.3  13.8  17.2  13.8  20.7  6.9  62.1 | p(abc) =  0.620  0.640  0.421  0.899  0.568  0.342  0.390  0.712  0.052  0.053  0.952  0.634 |

Note. Dopaminergic (DA) receptor antagonist medications include quetiapine, risperidone, olanzapine, ziprasidone, and prochlorperazine. DA receptor agonists include bupropion, amphetamines, dextroamphetamines, and methylphenidate. All participants prescribed acute-acting psychotropic medications refrained from taking that medication on the day of the scan and at least two hours afterwards. All scans occurred between the hours of 1300-1900. Benzodiazepine medication types, including alprazolam, diazepam, and lorazepam, did not differ between drug groups, *p*(abc) = 0.527. Additionally, stimulant medication types, including dextroamphetamine/levoamphetamine, phentermine, methylphenidate, and lisdexamfetamine, did not differ between drug groups, *p*(abc) = 0.440.

\*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; Na=21, Nb=18, Nc=15.

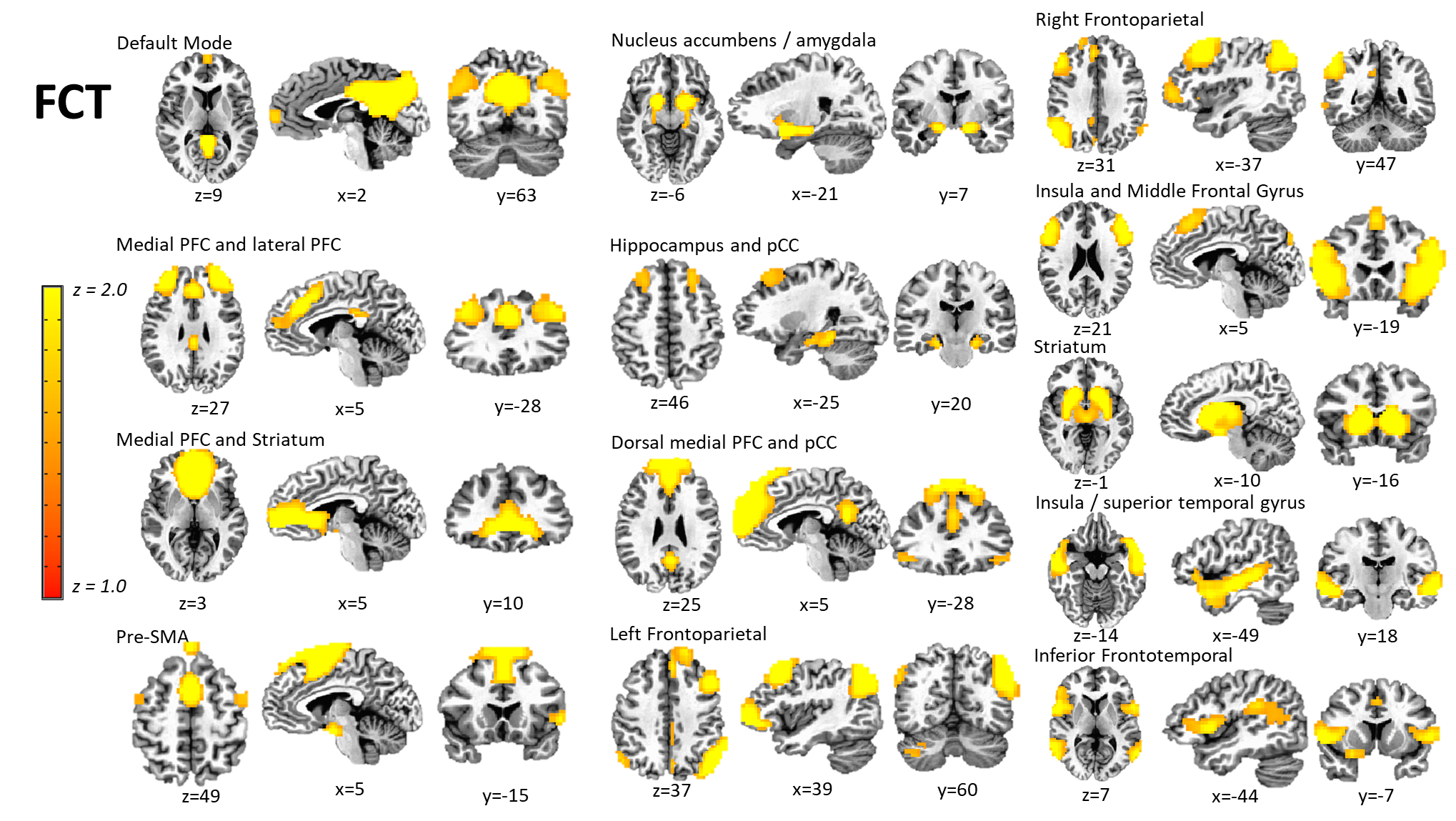


**Supplemental Figure 1**. Participant enrollment through the study design



**Supplemental Figure 2.** Fear Conditioning, Fear Extinction, and Fear Recall Task Structure. On Day 1 of scanning, participants completed the fear conditioning and fear extinction task in fMRI. The unconditioned stimulus (US) was an electric shock, which participants calibrated to an intensity level of 7/10 on a Likert scale. Conditioned stimuli consisted of triangles and circles, each displayed for 3s with a jittered inter-trial interval of 2-6s, and counterbalanced across participants. An initial baseline phase consisted of 6 presentations of each stimulus with no UCS onsets. The task then alternated between acquisition and extinction phases for 156 trials, with two presentations of each phase. The acquisition phase presented each CS 18 times, with a shock occurring 2.5s following CS+ presentation with a 50% reinforcement schedule. The extinction phase presented each stimulus 18 times and no shocks occurred. Participants returned for a recall test 24 hours later using a similar procedure and task design. Participants first recalibrated the US intensity again to a 7 on a 10 point Likert scale. The recall task alternated between acquisition and extinction contexts, in a pseudorandom order for a total of 6 context repetitions, and presented two CS+ and CS- stimuli per context presentation (i.e., 12 total CS+ and CS- repetitions). Each stimulus was again presented for 3s with a jittered inter-trial interval of 2-6s. After the first run of the task, participants then received a single uncued US presentation (i.e., reinstatement), and then participants completed again the identical task as implemented in the first run.

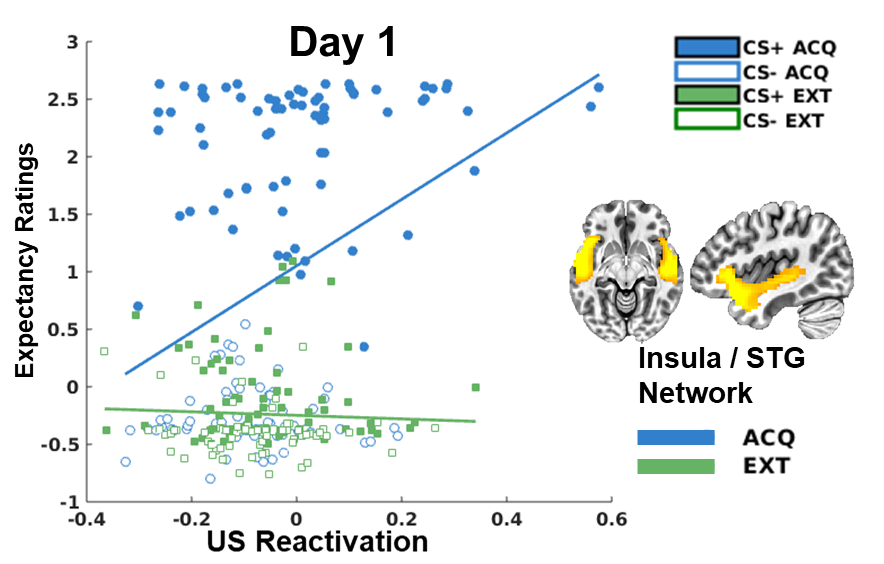
**Supplemental Figure 3.** Networks of interest from the Independent Component Analysis (ICA) conducted on Day 1 and Day 2 fear conditioning, extinction, and recall task data. The ICA used a model order of 35, and 13 of these components were selected as networks of interest after removing components attributed to artifact (e.g., CSF, head motion) and networks of non-interest (e.g., visual and motor networks). FCT = fear conditioning tasks.

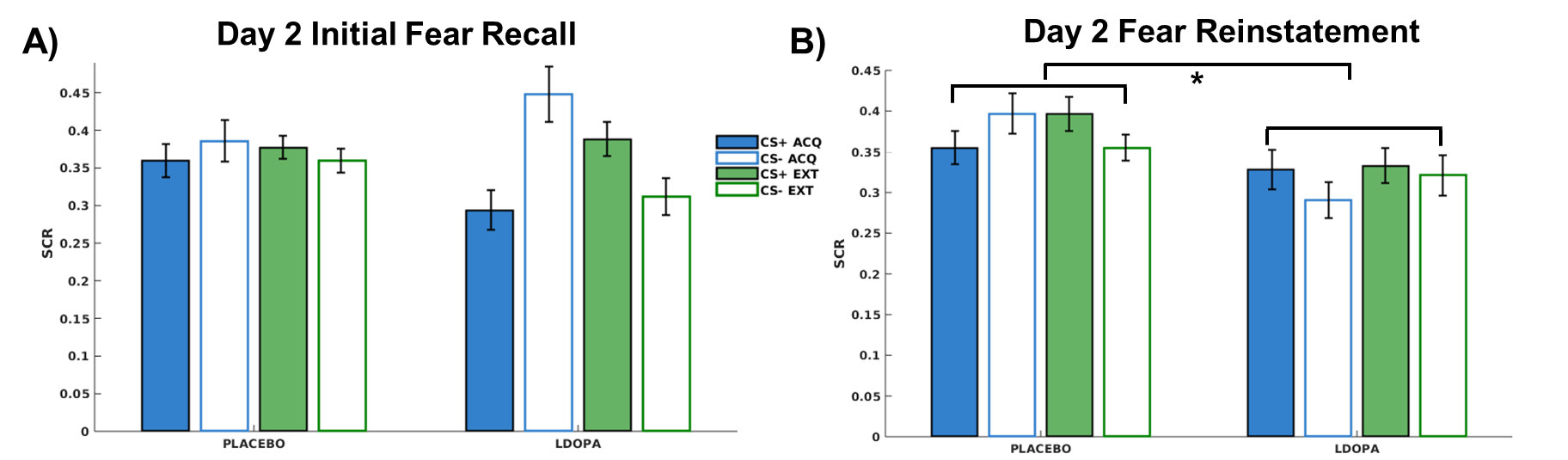


A picture containing text, screenshot, diagram, parallel

Description automatically generated

**Supplemental Figure 4**. Skin conductance response (SCR) data for each alternation of the acquisition and extinction phases. Data are separated into early, middle, and late blocks within each phase. B-D).

**Supplemental Figure 5**. Skin conductance response (SCR) data for each alternation of the acquisition and extinction phases. Data are separated into early, middle, and late blocks within each phase. B-D). The specificity for this association within the acquisition context is likely due to the restricted range of expectancies in the extinction context; that is, expectancies were consistently low towards both CS+ and CS- in the extinction context.



**Supplemental Figure 6**.Skin conductance response (SCR) data from the first block of the initial fear recall test (A) and following Fear Reinstatement (B) on Day 2.

Supplemental References

1. Haaker J, Gaburro S, Sah A, Gartmann N, Lonsdorf TB, Meier K, *et al.* (2013): Single dose of l-dopa makes extinction memories context-independent and prevents the return of fear. *Proc Natl Acad Sci* 110: E2428–E2436.

2. Raij T, Nummenmaa A, Marin M-F, Porter D, Furtak S, Setsompop K, Milad MR (2018): Prefrontal Cortex Stimulation Enhances Fear Extinction Memory in Humans. *Biol Psychiatry* 84: 129–137.

3. Garfinkel SN, Abelson JL, King AP, Sripada RK, Wang X, Gaines LM, Liberzon I (2014): Impaired Contextual Modulation of Memories in PTSD: An fMRI and Psychophysiological Study of Extinction Retention and Fear Renewal. *J Neurosci* 34: 13435–13443.

4. Bach DR (2014): A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis. *Biol Psychol* 103: 63–68.

5. Bach DR, Flandin G, Friston KJ, Dolan RJ (2009): Time-series analysis for rapid event-related skin conductance responses. *J Neurosci Methods* 184: 224–234.

6. Bach DR, Flandin G, Friston KJ, Dolan RJ (2010): Modelling event-related skin conductance responses. *Int J Psychophysiol Off J Int Organ Psychophysiol* 75: 349–356.

7. Bach DR, Friston KJ, Dolan RJ (2013): An improved algorithm for model-based analysis of evoked skin conductance responses. *Biol Psychol* 94: 490–497.

8. Bach DR, Friston KJ (2013): Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology* 50: 15–22.

9. Gerster S, Namer B, Elam M, Bach DR (2017): Testing a linear time invariant model for skin conductance responses by intraneural recording and stimulation. *Psychophysiology*. https://doi.org/10.1111/psyp.12986

10. Staib M, Castegnetti G, Bach DR (2015): Optimising a model-based approach to inferring fear learning from skin conductance responses. *J Neurosci Methods* 255: 131–138.

11. Bach DR, Tzovara A, Vunder J (2017): Blocking human fear memory with the matrix metalloproteinase inhibitor doxycycline. *Mol Psychiatry*. https://doi.org/10.1038/mp.2017.65

12. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, *et al.* (2018): Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* 167: 104–120.

13. Yu M, Linn KA, Cook PA, Phillips ML, McInnis M, Fava M, *et al.* (2018): Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum Brain Mapp* 39: 4213–4227.

14. Molapour T, Golkar A, Navarrete CD, Haaker J, Olsson A (2015): Neural correlates of biased social fear learning and interaction in an intergroup context. *NeuroImage* 121: 171–183.

15. Morriss J, Hoare S, van Reekum CM (2018): It’s time: A commentary on fear extinction in the human brain using fMRI. *Neurosci Biobehav Rev* 94: 321–322.