**Supplemental Material**

**Methods**

**Participants**

Comorbid diagnoses were generalized anxiety disorder (26.7%, n=24), social anxiety disorder (26.7%, n=24), major depressive disorder (26.7%, n=24), panic disorder (21.1%, n=19), persistent depressive disorder (15.6%, n=14), specific phobia (14.4%, n=13), posttraumatic stress disorder (13.3%, n=12), eating disorder (5.6%, n=5), alcohol abuse (2.2%, n=2), obsessive compulsive disorder (2.2%, n=2), and acute adjustment disorder (1.1%, n=1).

All participants were between 18 and 65 years of age, free of major medical or neurologic illness as confirmed by a Board-Certified physician. Exclusion criteria for all participants were as follows: inability to provide consent; current or history of psychotic symptoms, bipolar disorder, traumatic brain injury, mental retardation, pervasive developmental disorder, or dementia; active suicidal ideation (within 6 months of study entry); current alcohol or substance dependence (within 6 months of study entry); treatment (e.g., pharmacotherapy, psychotherapy) at the time of study entry and during the study; previous trial of Cognitive Behavioral Therapy; and contraindications to magnetic resonance imaging (e.g., pregnancy, non-removable ferrous objects, claustophobia). Healthy participants could not have current or lifetime history of an Axis I disorder. None of the participants tested positive on a urine toxicology screen before scans.

Participants were recruited from the Mood and Anxiety Clinic at the University of Illinois at Chicago via referrals and in the community through flyers and internet advertisements. Interested participants completed a phone screen followed by a psychiatric evaluation during which time they reviewed the consent form as approved by the local Institutional Review Board at the University of Illinois at Chicago. After attaining consent, all participants completed the Structured Clinical Interview (i.e., SCID-5) by a non-treating trained staff clinician. Eligibility for treatment was based on the Structured Clinical Interview(First et al., 2015) relevant clinical assessments by a non-treating trained staff clinician, and Best-Estimate/Consensus Panel of study staff members (e.g., clinical psychologist, psychiatrist, clinical assessor). A non-treating trained staff clinician assessed symptom severity before and immediately after CBT was completed.

**Open trial CBT vs. randomized-to-CBT trial**

Data was collected across two studies. One comprised an open trial of 12 weeks of cognitive behavioral therapy (CBT) with patients diagnosed with social anxiety disorder (n=32), major depressive disorder (n=13), generalized anxiety disorder (n=9), or panic disorder (n=2). The other involved a clinical trial where patients were randomized to 12 weeks of CBT or medication. In the clinical trial study, analysis was limited to patients who completed the CBT arm, which consisted of patients diagnosed with generalized anxiety disorder (n=16), social anxiety disorder (n=6), major depressive disorder (n=5), panic disorder (n=3), posttraumatic stress disorder (n=3), or persistent depressive disorder (n=1).

Independent t-tests revealed patients from the open trial study were more socially anxious (LSAS: 72.36 ± 22.11) than randomized-to-CBT patients (LSAS: 59.03 ± 25.08) [*t*(88)=2.63, p=0.010] but endorsed less symptom severity (DASS: 26.16 ± 11.63) than the randomized sample (DASS: 30.91 ± 6.37) [*t*(88)=2.187, p=0.031]. Yet, patients from the open trial study were equivalent in general anxiety level (HAMA: 15.25 ± 8.45) compared to randomized-to-CBT patients (HAMA: 16.82 ± 5.00) [*t*(88)=0.98, p=0.33] similar in depression level (HAMD: 10.09 ± 6.11) compared to randomized-to-CBT patients (HAMD: 12.18 ± 3.73) [*t*(88)=1.79, p=0.08], and equivalent in worry level (PSWQ: 63.43 ± 8.68) compared to randomized-to-CBT patients (PSWQ: 65.56 ± 7.72) [*t*(88)=1.17, p=0.24].

Regarding demographic characteristics, open trial study patients were similar in age (25.18 ± 6.37) compared to randomized-to-CBT patients (27.17 ± 8.00) [*t*(88)=1.31, p=0.19] and gender frequency in the open trial study [female (n=41, 73.2%); male (n=15, 26.8%)] was comparable to the randomized sample [female (n=25, 73.5%); male (n=9, 26.5%)] [χ2(1)=0.001, p=0.97]. However, open trial study patients reported having less education in years (15.11 ± 2.04) relative to randomized-to-CBT patients (16.79 ± 3.34) [*t*(88)=2.97, p=0.011]. Studies were comparable as to ethnicity [χ2(1)=1.83, p=0.17] but differed in race [χ2(4)=9.81, p=0.04].

**fMRI Task**

The emotion face matching task comprised photographs from a validated set of face stimuli (Gur *et al.* 2002) presented in a block-design. Faces consisted of 3 angry, 3 fearful, and 3 happy blocks of trials[[1]](#footnote-1), interspersed with shape-matching blocks (triangles, circles, squares) as a sensorimotor control condition. Right-handed button press was used to match facial expressions and shapes. Each block lasted 20 seconds and consisted of 4 back-to-back 5-second trials.

**fMRI Data Acquisition and Preprocessing**

Functional imaging was performed with blood-oxygen-level-dependent(BOLD) sensitive whole-brain fMRI on a 3.0 Tesla GE Signa System (General Electric; Milwaukee, WI) with an 8-channel head coil. Functional data were acquired using gradient-echo echo planar imaging sequence with the following parameters: TR = 2s, TE = minFull [~25 ms], flip angle = 90º, FOV = 22 x 22 cm2, acquisition matrix 64 x 64, 3-mm slice thickness, 44 axial slices, 180 volumes per run. For anatomical localization, a high-resolution, T1-weighted volumetric anatomical scan was acquired.

Data from all participants were examined for movement during the scan and participants with excessive motion (movement greater than 3 mm in any direction or 3 degrees of displacement across each functional run) were excluded and the first 4 volumes from each run were discarded to allow for T1 equilibration effects. Conventional preprocessing steps were used in Statistical Parametric Mapping (SPM12) software package (Wellcome Trust Centre for Neuroimaging, London [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Briefly, images were temporally corrected to account for differences in slice time collection, spatially realigned to the first image of the first run, normalized to a Montreal Neurological Institute (MNI) template, and smoothed with an 8 mm isotropic Gaussian kernel.

A box-car model was used to analyze the following contrasts of interest: angry (>shapes), fearful (>shapes), and happy (>shapes).

**Decision Tree Analysis**

The exhaustive Pearson chi-squared automatic interaction detection (CHAID) algorithm determines how variables best combine to predict a binary outcome based on ‘if-then’ logic by partitioning each variable into mutually exclusive subsets based on the homogeneity of data. Accordingly, it generates all possible cross tabulations for each predictor and starts with a binary split (i.e., root node) into two or more homogeneous sets until no further splits can be made (Kass, 1980). Subsequent splits from root node may be parent nodes if they precede other splits (i.e., child/terminal nodes) or be child nodes themselves if no other splits are made. Consistent with a prior MRI study (Wu *et al.* 2017), the CHAID algorithm was performed with >25 cases per parent node and >20 cases per child node.

**Results**

**Open trial CBT vs. randomized-to-CBT trial**

A mixed 2 study sample (open trial, randomized-to-CBT trial) by 3 primary outcome measure (HAMA, HAMD, DASS) by 2 time (Pre-CBT, Post-CBT) repeated measures Multivariate Analysis of Variance (MANOVA) revealed a main effect of time [F(1, 88)=276.76, p<0.001] and measure [F(2, 176)=226.42, p<0.001]. There was no main effect for study sample [F(1, 88)=1.36, p=0.247] though there was a time x measure x study sample interaction [F(2, 176)=7.47, p=0.002] and time x study sample interaction [F(1, 88)=4.53, p=0.036]. There was also a measure x time interaction [F(2, 176)=87.15, p<0.001] but no measure x study sample interaction [F(2, 176)=0.05, p=0.912].

To evaluate the main effect of time, measures and study sample were collapsed. A paired t-test showed symptom severity across measures were higher before CBT (18.22 ± 6.67) than after completing CBT (6.87 ± 5.53) [t(89)=16.30, p<0.001]. To evaluate the main effect of measure, we collapsed across time and study sample. Paired t-tests showed all patients exhibited more anxiety indexed with HAMA (10.77 ± 5.28) than depression assessed with HAMD (10.05 ± 4.86) [t(89)=5.11, p<0.001]. However, self-reported symptom severity via DASS was higher (19.30 ± 7.94) than HAMA [t(89)=13.82, p<0.001] and HAMD [t(89)=14.89, p<0.001] across patients.

To examine the highest-order interaction, we performed independent t-tests with study sample as the between-subjects group for each measure before and after CBT. As reported above, patients from the open trial study at baseline were more socially anxious than the randomized-to-CBT patients but self-reported less symptom severity than the randomized sample. Also, patients in both samples had analogous levels of general anxiety, depression, and worry. After completing CBT, independent t-tests revealed patients from the open trial study were similar in general anxiety level (HAMA: 5.43 ± 5.31) compared to the randomized-to-CBT patients (HAMA: 6.12 ± 4.77) [*t*(88)=0.62, p=0.54], similar in depression level (HAMD: 3.91 ± 4.39) relative to randomized-to-CBT patients (HAMD: 4.82 ± 3.94) [*t*(88)=0.99, p=0.32], and similar in self-reported symptom severity (DASS: 11.48 ± 9.36) compared to randomized-to-CBT patients (DASS: 9.29 ± 8.71) [*t*(88)=1.10, p=0.27].

Paired t-tests results within each study sample was analogous to the main effect of time finding across primary measures. Specifically, in the open trial study, anxiety at baseline (15.25 ± 8.45) decreased after CBT (5.43 ± 5.30) [HAMA; t(55)=10.01, p<0.001], depression level at baseline (10.09 ± 6.11) decreased after CBT (3.91 ± 4.39) [HAMD; t(55)=7.73, p<0.001], and self-reported symptom severity at baseline (26.16 ± 11.64) decreased after CBT (11.48 ± 9.36) [DASS; t(55)=9.41, p<0.001]. Comparable results were observed in the randomized-to-CBT sample as anxiety at baseline (16.82 ± 5.00) decreased after CBT (6.12 ± 4.77) [HAMA; t(33)=9.93, p<0.001], depression level at baseline (12.18 ± 3.73) decreased after CBT (4.82 ± 3.94) [HAMD; t(33)=8.26, p<0.001], and self-reported symptom severity at baseline (30.91 ± 6.37) decreased after CBT (9.29 ± 8.72) [DASS; t(33)=14.71, p<0.001].

When the same MANOVA was performed for secondary measures, there was a main effect of time [F(1, 88)=154.68, p<0.001] and a non-significant trend towards a main effect of measure [F(1, 88)=3.92, p=0.05]. There was no main effect of study sample [F(1, 88)=1.39, p=0.24] and no measure x time x study sample interaction [F(1, 88)=3.34, p=0.07]. However, there was a measure x study sample interaction [F(1, 88)=8.65, p=0.004] and measure x time interaction [F(1, 88)=45.81, p<0.001].

To evaluate the main effect of time, we collapsed across measure and study sample. Paired-tests results revealed a composite social anxiety and worry score was higher at baseline (65.78 ± 13.89) than post-CBT (47.32 ± 18.26) [t(89)=13.10, p<0.001]. To examine the measure x study sample interaction, we collapsed across time. Paired t-tests in the open trial sample showed no difference between social anxiety (LSAS; 50.00 ± 22.22) and worry (PSWQ; 56.95 ± 9.30) [t(55)=0.78, p=0.43]. Conversely, in the randomized-to-CBT sample, social anxiety level (LSAS; 48.95 ± 24.68) was less than worry (PSWQ, 59.44 ± 9.82) [t(33)=3.08, p=0.004]. To evaluate the measure x time interaction, we collapsed across study sample. Across patients, paired t-tests showed social anxiety level was higher at baseline (LSAS; 67.32 ± 24.04) than post-CBT (LSAS; 43.09 ± 26.69) [t(89)=12.12, p<0.001]. Additionally, worry level was higher at baseline (PSWQ; 64.23 ± 8.35) than post-CBT (PSWQ; 51.56 ± 12.87) [t(89)=11.57, p<0.001].

**Task performance**

Accuracy for the 3 (Group: responder, non-responder, healthy control) x 4 (Stimulus Type: angry, fearful, happy, shapes) x 2 (Time: Week 0 and Week 12) mixed repeated measures ANOVA revealed a main effect of stimulus type (*F*(3,396)=96.94, p<0.001) but no main effects of group, time, or interactions (lowest p=0.522). To evaluate the main effect of stimulus type, accuracy was collapsed across group and time. Follow-up paired t-tests showed participants had greater accuracy to happy faces (96.3% ± 0.08%) compared to fearful faces (93.3% ± 0.09%), angry faces (84.9% ± 0.11%), and shapes (91.6% ± 0.06%) (all p’s<0.001). Accuracy was higher for fear faces relative to angry faces (p<0.001) and shapes (p<0.003). Accuracy was also higher for happy faces compared to angry faces (p<0.001).

When the same analysis was performed for RT for accurate trials in milliseconds (ms), there were main effects for stimulus type (*F*(3,396)=277.02, p<0.001) and time (*F*(1,396)=23.37, p<0.001) moderated by a stimulus type x time interaction (*F*(3,396)=7.10, p<0.001). There was no main effect for group or other interactions (lowest p=0.41).

To examine the main effect of time, we collapsed across stimulus type and group. The paired t-test revealed RT was slower at baseline (1317.52 ± 283.83 ms) than after 12 weeks (1205.93 ± 272.94 ms) (p<0.001). To evaluate the main effect of stimulus type, we collapsed time and group. Paired t-tests showed RT for shapes (903.62 ± 171.63 ms) was faster than for angry faces (1418.51 ± 357.43 ms), fearful faces (1474.12 ± 326.59 ms), and happy faces (1330.35 ± 281.72 ms) (all p’s<0.001). Also, RT was faster for happy faces compared to fearful faces (p<0.001) and angry faces (p=0.002) and RT was faster for angry faces than fearful faces (p=0.042).

To examine the stimulus type x time interaction, follow-up paired t-tests were performed for stimuli type before and after 12 weeks. Across participants, matching angry faces at baseline (1473.79 ± 411.31 ms), fearful faces (1527.55 ± 422.14 ms), and happy faces (1358.87 ± 309.91 ms) was slower than after 12 weeks (angry: 1314.96 ± 411.19 ms; fearful: 1377.10 ± 363.29 ms; happy 1243.45 ± 335.44 ms) (all p’s<0.001). Yet, RT for shapes was similar before (909.85 ± 205.72 ms) and after 12 weeks (888.19 ± 188.87 ms) (p=0.23).

**Classified treatment-outcome subgroups following treatment**

For comprehensiveness we evaluated laterality even though right superior occipital gyrus (SOG) was the classifier. A 3 (Group: LO subgroup, HO subgroup, healthy controls) x 2 (Hemisphere: left, right) x 2 (Time: Week 0, Week 12) mixed ANOVA showed a main effect of group [*F*(2, 132)=23.35, p=0.001] and hemisphere [*F*(1, 132)=8.24, p=0.005] but no main effect of time [*F*(1, 132)=0.04, p=0.84]. There was a group x time interaction [*F*(2, 132)=11.77, p=0.001] and group x hemisphere interaction [*F*(2, 132)=4.58, p=0.012] but no other interactions (lowest p=0.11).

For the main effect of group, hemispheres were collapsed; and a one-way ANOVA revealed group effects over time (all p’s<0.05). Post-hoc Bonferroni corrected results showed the LO subgroup exhibited less activation at baseline than both the HO subgroup (p=0.001) and control group (p=0.001). The HO subgroup also demonstrated more activation at baseline than the LO (p=0.001) and control (p=0.002) groups. Following CBT, both subgroups had similar SOG activity as controls (lowest p=0.28). Between patient subgroups, the LO subgroup continued to show relatively less activation than the HO subgroup even after completing CBT (p=0.035).

For the main effect of hemisphere, time and group were collapsed. Paired t-tests revealed more activation in right SOG compared to left SOG (*F*(2, 134)=3.24, p=0.001).

For the group x hemisphere interaction, time was collapsed. A one-way ANOVA showed group effects for left (p=0.001) and right hemispheres (p=0.001). Post-hoc Bonferroni corrected results revealed the LO subgroup exhibited less left SOG activation at baseline than the HO subgroup (p=0.001) and controls (p=0.001). However, the HO subgroup and controls showed analogous activation (p=0.36). For right hemisphere, the LO subgroup demonstrated less right SOG activation at baseline than the HO subgroup (p=0.001) and controls (p=0.001). Additionally, the HO subgroup exhibited more activity in right SOG than controls (p=0.001).

For group x time interaction, hemispheres were collapsed; paired t-tests revealed the LO subgroup exhibited an increase in bilateral SOG activity following CBT [*t*(35)=3.60, p=0.001] whereas the in the HO subgroup, bilateral SOG activity decreased following CBT [*t*(53)=3.93, p=0.001].

**SFigure 1.** **Illustration of *a priori* Regions of Interest**

|  |
| --- |
| A. |
| B. |
| C. |

**A)** Frontal ROIs. Red – Inferior frontal gyrus, opercular portion. Green – Inferior frontal gyrus, orbital portion. Dark Blue – Inferior frontal gyrus, triangular portion. Yellow – Frontal medial orbital gyrus. Light Blue – Middle frontal gyrus (i.e. dorsolateral prefrontal cortex). Gold – Medial superior frontal gyrus. **B)** Temporal and occipital ROIs. Red – Occipital cortex, inferior portion. Green – Occipital cortex, middle portion. Dark blue – Occipital cortex, superior portion. Light blue – Middle temporal gyrus. Gold – Superior temporal gyrus. Yellow – Inferior temporal gyrus. **C)** Subcortical ROIs. Red – Amygdala. Green – Anterior cingulum. Dark blue – Hippocampus. Yellow – Insula.

1. A sub-sample of patients (n=72) and healthy controls (n=38) also completed a modified version of the task that included sad faces. However, restricting analysis to participants who completed the version with angry, fearful, happy, and sad faces would reduce the sample size. Therefore, we elected to perform analysis on the larger sample who completed the original task. [↑](#footnote-ref-1)