Task-based neural correlates of self-focused attention before and after cognitive behavioral therapy

Emily M. Iannazzi

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Angela Fang

Lori Zoellner

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Task-based neural correlates of self-focused attention before and after cognitive behavioral therapy

Emily M. Iannazzi

Chair of the Supervisory Committee:

Angela Fang

Psychology

Self-focused attention (SFA), a form of self-referential processing, is maladaptive in various psychiatric disorders and may be associated with poor treatment response. This study examined SFA in individuals with social anxiety disorder (SAD) and body dysmorphic disorder (BDD), testing the hypothesis that SFA is associated with hyperactivity within default network (DN) regions and with treatment response during cognitive-behavioral therapy. Participants included 30 patients with primary SAD or BDD and 28 healthy controls, who displayed above average and below average scores (respectively) on the Public Self-Consciousness Scale, which measured trait SFA. SFA was also measured by a self-referential encoding task, which yielded both behavioral reaction time measures and task-related fMRI measures of SFA. Results indicated significantly longer reaction times at pre-treatment for self vs. other trials in patients

compared to controls, with patients showing notable improvement post-treatment. Neuroimaging revealed greater activation in DN regions, including the medial prefrontal cortex, during self vs. other trials in all participants; however, there were no significant group differences at pre- or post-treatment, nor in the changes from pre- to post-treatment. Neural measures of SFA were significantly associated with treatment response, whereas behavioral measures were not. These findings suggest that activity in DN regions may serve as a transdiagnostic biomarker of maladaptive SFA that is associated with treatment response.

Task-based neural correlates of self-focused attention before and after cognitive behavioral therapy

Self-focused attention (SFA) is a form of self-referential processing which refers to a tendency to reflect on internal thoughts, feelings, beliefs, and physical states (Ingram, 1990). In psychiatric disorders, SFA is excessive and maladaptive. Related forms of aberrant self-referential processing, such as worry and rumination, have been linked to interference with corrective learning (Reilly et al., 2018), symptom relapse (LeMoult et al., 2017), and treatment non-response (Jones, Siegle, & Thase, 2008). In depression in particular, aberrant self-referential processing may persist in remission and pose a latent risk factor for future recurrence of depression (Allison et al., 2023). Although SFA has been strongly implicated across many disorders, it has been the most well-studied in depression and its "transdiagnostic" nature remains unclear. In the current study, we examined SFA in two pathologies characterized by maladaptive levels of SFA, social anxiety disorder (SAD) and body dysmorphic disorder (BDD), and tested the hypothesis that neural correlates of SFA may predict improved treatment response during cognitive behavioral therapy (CBT).

SFA may be measured across behavioral, self-report, and neural levels of analysis. The most commonly-used behavioral measure of SFA is the self-referent encoding task (SRET). The SRET is designed to measure reaction times when making judgments about whether trait adjectives are descriptive of oneself or not (Kelley et al., 2002). Individuals with anxiety and depression tend to endorse more negative than positive words as self-descriptive, and display longer reaction times for negative (versus positive) self judgments (Collins & Winer, 2023; Hitchcock et al., 2023). In terms of self-report measures of SFA, the Self-Consciousness Scales Revised (SCS-R; Scheier & Carver, 1985) is a well-validated measure of public and private

components of trait SFA in both non-clinical and clinical populations. An earlier meta-analysis showed that public SFA (aspects of self that involve others) was specifically linked to anxiety, especially social anxiety disorder (SAD), whereas private SFA (aspects of self that do not involve others) was specifically linked to depression (Mor & Winquist, 2002). Thus, we utilized the SRET and Public SCS-R measures in our study as behavioral and self-report measures of SFA, respectively.

On the neural level, several neuroimaging studies have found associations between SFA and anatomical components of the default mode network (DN), such as the medial prefrontal cortex, temporoparietal junction, and temporal pole (Abraham et al., 2013; Boehme et al., 2015; Freton et al., 2014). Task-based fMRI studies probing self-referential processing have been conducted across various analogue and clinical samples, mainly in depression and SAD (Dixon et al., 2022; Grimm et al., 2009; Jamieson et al., 2023). Two studies have examined neural correlates of the SRET in clinical SAD samples, largely showing similar patterns of DN activation during self-referential processing in patients with SAD and healthy controls and no group differences that survive multiple comparisons (Dixon et al., 2022; Jamieson et al., 2023). This may reflect that actual differences are small and studies were underpowered to detect such differences, or that heterogeneity in SAD, in relation to self-referential processing, obscured real differences, or both. We address the issue of heterogeneity in the current study by sampling patients on the basis of displaying maladaptive trait SFA on the Public SCS-R measure.

Previous studies suggest that SFA can be manipulated by various interventions and that effects of the manipulations correlate with symptom improvement. For example, in one study, patients with SAD, compared to healthy controls, displayed a pattern of neural hyperactivity in bilateral temporoparietal regions and in dorsal components of the posterior cingulate cortex when

viewing aversive compared to neutral pictures–an effect which normalized after training them to regulate their emotions by modifying their personal relevance to the situation (called "self-focused reappraisal") (Gaebler et al., 2018). In another study, self-focused thoughts were assessed in anticipation of a social stress task before and after exposure therapy (Hofmann, 2000). Negative self-focused thoughts were significantly reduced after treatment, and changes in negative self-focused thoughts were significantly associated with symptom improvement. In depression, antidepressant treatment has also been shown to reduce negative self-referential processing (DiSimplicio et al., 2012; Komulainen et al., 2018). These studies suggest that SFA may be an indicator of therapeutic change during psychological and pharmacological treatment. Since no reliable demographic or clinical predictors of treatment response have been identified for psychiatric disorders (Schneider et al., 2015), research examining alternative indicators of therapeutic success have important implications for treatment selection and treatment planning decisions.

Neural correlates of SFA may be a candidate biomarker of CBT. Our previous analysis in the same sample showed that pre-treatment resting state functional connectivity within the DN was associated with CBT response (Fang et al., 2024). In the current study, we extended this work as well as previous work on self-referential processing in non-depressed samples by focusing on a transdiagnostic anxious sample (SAD and BDD) selected for maladaptive levels of SFA and by focusing on task-related neural correlates. First, we hypothesized that patients with SAD and BDD with maladaptive SFA would display greater regional brain activation of DN regions (specifically the medial prefrontal cortex, temporoparietal junction, and posterior cingulate) during self-referential processing, compared to non-self-focused healthy controls, and that they would also exhibit longer response latencies when making self-relevant versus other-

relevant judgments. Second, we hypothesized that the pre-treatment neural correlates of SFA would be associated with CBT response, and that this association would be stronger than that of behavioral measures of SFA based on response latencies. Finally, given previous reports that CBT appears to normalize SFA in SAD patients (Hofmann et al., 2004), we examined whether CBT reduced SFA across behavioral and neural measures of analysis, and whether changes would correlate with clinical improvement.

Method

Participants

Participants were 28 healthy controls and 30 participants diagnosed with primary SAD (N=17) or BDD (N=13) based on the SCID-5-RV (First et al., 2015). Participants were recruited through online and print advertisement, clinician referral, and self-referral to specialized clinics in anxiety and obsessive-compulsive related disorders at a large metropolitan hospital. Participants were matched on age, sex, race, ethnicity, and education across patients and controls. Patients with BDD and SAD were also matched on these variables except for age, with the SAD participants being significantly older than those with BDD (SAD: M=27.00, SD=4.41; BDD: M=23.87, SD=2.13; $t = -2.48$, $p = 0.02$). Further description of this sample can be found in Supplementary Table S1 and Fang et al. (2022).

Inclusion Criteria

Shared inclusion criteria for both groups consisted of: 1) Men and women aged 18-45, and 2) Right hand dominance (per Edinburgh Handedness Inventory) to ensure uniform hemispheric dominance for interpretation of neuroimaging data. Additional inclusion criteria for clinical participants were: 1) treatment-seeking individuals presenting with at least moderate levels of socially anxious or body dysmorphic symptoms (e.g., meeting cutoff score of ≥ 50 on

the Liebowitz Social Anxiety Scale for socially anxious concerns, and score of ≥ 20 on the Yale-Brown Obsessive-Compulsive Scale Modified for BDD for body dysmorphic concerns), and 2) a score ≥ 1 SD of the normative mean on the SCS-R (≥ 18) (Scheier & Carver, 1985). Additional inclusion criteria for healthy controls were: 1) a score of ≤ 1 SD of the normative mean on the SCS-R (≤ 9) (Scheier & Carver, 1985), and 2) no lifetime history of psychiatric disorders, based on the SCID-5-RV.

Exclusion Criteria

Exclusion criteria for both groups consisted of: 1) Positive screen for MR contraindicators, 2) History of head injury, neurological disorder, or neurosurgical procedure, 3) Active suicidal or homicidal ideation, or any features requiring a higher level of care, 4) Current or past manic/ hypomanic episode or psychotic symptoms, 5) Active alcohol or substance dependence in the past year, 6) Current CBT and/or formal mindfulness/meditation training, 7) History of more than 10 sessions of CBT and formal mindfulness/meditation training, and 8) Current use of psychotropic medications (except for clinical participants who were allowed to take stable psychotropic medications for at least 2 months prior to the first study visit). Clinical participants agreed to not begin other treatments or medications during the study.

Psychometric Measures

Structured Clinical Interview for DSM-5 Research Version (SCID-5-RV)

The SCID-5 (First et al., 2015) was used to assess BDD and SAD according to the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). This semi-structured interview was administered by a licensed psychologist who served as the study therapist (A.F.) and assessed inclusion/exclusion criteria for both clinical participants and healthy controls.

Liebowitz Social Anxiety Scale (LSAS)

The LSAS (Liebowitz, 2013) is an interviewer-rated instrument that was administered to participants with primary SAD to evaluate past week severity of social anxiety symptoms. Responses are loaded onto two subscales for fear and avoidance in a range of social situations and subscale scores are summed to yield a total severity score.

Yale-Brown Obsessive-Compulsive Scale Modified for Body Dysmorphic Disorder (BDD-YBOCS)

The BDD-YBOCS (Phillips et al., 1997) is an interviewer-rated instrument that assesses past week severity of BDD symptoms based on the frequency and interference of BDD-related obsessions and compulsions.

Brown Assessment of Beliefs Scale (BABS)

The BABS (Eisen et al., 1998) is a 7-item interviewer-rated measure of insight and delusionality in the past week. Items are anchored to a principal belief related to their SAD/BDD symptoms, such as "I am awkward" or "I look horrible." The BABS is required to calculate the BDD-YBOCS total score and was used for all clinical participants to examine level of insight.

Self-Consciousness Scale-Revised (SCS-R)

The Public subscale of the SCS-R (Scheier & Carver, 1985) was selected because it is a widely used and reliable measure of trait SFA (Mor & Winquist, 2002). This scale was used during the pre-screening process to determine eligibility of having elevated levels of trait SFA ($\geq$18) for clinical participants and low levels ($\leq$9) for healthy controls.

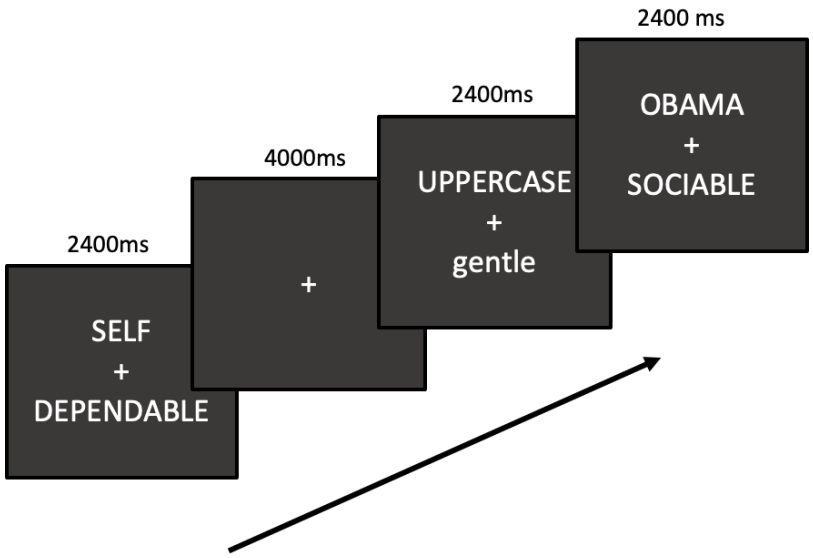
Self-Referent Encoding Task

Participants completed the Self-Referent Encoding Task (SRET; Kelley et al., 2002) while undergoing functional magnetic resonance imaging (fMRI) at pre- and post-treatment time

points. Participants in the control group also underwent the same procedure with pre and post time points occurring 12 weeks apart. The SRET is an event-related fMRI task that asks participants to make judgments about trait adjectives relating to "self" or "other" (Figure 1). A total of 270 unique adjectives were selected (Appendix A) from a pool of normalized adjectives (Anderson, 1968). Lists were counterbalanced for word length, number of syllables, and valence. Behavioral and neural data were extracted from this task. Longer reaction times for self vs. other trials indicated greater SFA and represented a behavioral measure of SFA.

Figure 1

Event-related task design



Note. Participants were asked to make judgments about trait adjectives comprising one of three types: Self ("Does this adjective describe you?"), Other ("Does this adjective describe former U.S. President Barack Obama?"), and Uppercase ("Is this adjective printed in uppercase

letters?"). Each trial lasted 2500 ms and consisted of a "cue" word (Self, Obama, Uppercase) above a central fixation and a unique trait adjective (e.g., "POLITE") below a central fixation presented for 2000 ms.

Treatment

Participants in the clinical group received 12 50-minute weekly individual sessions of CBT provided by the same therapist (A.F.). The Hofmann & Otto (2008) CBT for SAD and Wilhelm et al. (2013) CBT for BDD treatment manuals were implemented for patients with primary SAD or BDD, respectively. Both treatments rely on core CBT principles and follow identical treatment components: psychoeducation (sessions 1-2), cognitive restructuring (sessions 3-4), introduction of exposure and in-vivo exposure practice (sessions 5-11), and relapse prevention (session 12). The therapist completed the Patient Exposure and Response Prevention Adherence Scale (PEAS; Simpson et al., 2010) to assess homework compliance between sessions 7-12. Treatment response was defined by a standardized continuous variable based on percent symptom reduction between pre- to post-treatment on the respective interviewer-rated symptom severity scale (LSAS for primary SAD or BDD-YBOCS for primary BDD).

Independent Evaluator Assessments

A doctoral-level independent evaluator (IE) was trained to achieve inter-rater reliability for clinical outcome measures in the study, which included the BDD-YBOCS, LSAS, and BABS. The IE attended weekly rater supervision to minimize bias and rater drift over the course of treatment. Intraclass correlations (ICC) for each measure were very high (ICCs > 0.99) (Shrout & Fleiss, 1979). Clinical participants met with the IE three times during the study (pre-

treatment, mid-treatment after treatment session 6, and post-treatment after treatment session 12). Healthy controls did not complete any IE assessments.

Procedure

Participants were first pre-screened by phone with the Public Self-Consciousness subscale of the Self-Consciousness Scales- Revised (SCS-R). Healthy controls who scored -1SD below the normative mean (≤9) and clinical participants who scored +1SD above the normative mean (≥18) on the Public SCS-R (Scheier & Carver, 1985) were invited for an in-person screening assessment. During the in-person screening visit, participants were assessed by a licensed psychologist specializing in anxiety and obsessive-compulsive related disorders (A.F.) for relevant psychiatric, medical, and neurological history using the SCID-5-RV. Eligible clinical participants completed IE assessments, which included the BDD-YBOCS (if primary BDD), LSAS (if primary SAD), and BABS, which were conducted at pre-, mid-, and post-treatment. Clinical participants then completed a pre-treatment MRI scan before undergoing individual CBT for 12 sessions with the study therapist. After the final CBT session, clinical participants returned for a post-treatment MRI scan within 7-10 days. Healthy controls were scanned twice at a 12-week interval to match the timeline for clinical participants but did not receive treatment.

MRI Acquisition and Preprocessing

Participants were scanned in a 3T Siemens Prisma scanner (Siemens Medical Systems, Iselin, NJ). Data preprocessing was done using fMRIPrep 20.2.0 (Esteban et al., 2019). All participants completed three functional runs of the SRET while undergoing fMRI using a T2*-weighted gradient echo sequence for blood oxygen level dependent (BOLD) contrast, as well as a T1-weighted high-resolution anatomical scan consisting of a rapid gradient-echo MEMPRAGE

sequence. Further details of data acquisition and preprocessing, including quality control and motion correction, can be found in the Supplementary Material.

Functional Neuroimaging Data Analysis

Subject and group-wise analyses were performed in the open-source python package Nilearn (Abraham et al., 2014). As all of our hypotheses were specifically related to DN functioning, we completed regions of interest (ROI) analyses in the DN. DN regions were defined using a previously validated Neurosynth-derived mask (Wang et al., 2020; Yarkoni et al., 2011) of the DN that included the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), and the bilateral temporoparietal junction (TPJ). The derived mask comprised 9,650 activations from 366 studies and was additionally constrained to the above-mentioned regions using the Harvard-Oxford probabilistic atlas (Desikan et al., 2006) at $p<0.25$. Additional information on mask creation and validation can be found elsewhere (Wang et al., 2020). Corrections within the ROI analyses were applied using $p_{FDR}<0.05$. We also conducted whole brain analyses for each hypothesis.

To examine group differences in regional brain activation during the SRET, we performed generalized linear models (GLM) based on the pre-treatment time point to estimate within-session, subject-specific parameter estimates to four modeled trial types: self, other, case, and fixation. Before computing the GLM, the three fMRI runs were concatenated with a run-wise intercept. An additional 32 head-motion parameters were included as additional regressors of no interest including the 6 rigid-body motion parameters, the average signal within the anatomically-derived CSF mask, the average signal anatomically-derived within the white matter mask, and the temporal derivatives and quadratic terms of each parameter (Satterthwaite et al., 2013). To mitigate measurement distortion across the multiple timepoints, we also masked the

model to constrain to just gray matter as determined by ICBM gray-matter mask (Fonov et al., 2009). The hemodynamic response function was modeled using a convolution of the experimental paradigm with a canonical hemodynamic response function. The contrast of interest, "self-other", reflected a contrast of the estimated self-trial parameters minus the estimated other-trial parameters.

To compare neural and behavioral measures of SFA in their association with treatment response, we conducted two separate linear regressions with extracted BOLD parameter estimates from self vs. other trials as the predictor in the neural model, and mean RT differences from self vs. other trials as the predictor in the behavioral model. Predictors were based on the pre-treatment time point. Treatment response was defined as percent symptom reduction from pre- to post-treatment based on the symptom severity scale for the primary disorder (e.g., BDD-YBOCS for primary BDD; LSAS for primary SAD). Since pre-treatment symptom severity was not a significant predictor of treatment response, it was not included in the models (Fang et al., 2024). Correlations between predicted and observed values for each model were computed and Fisher's Z test was used to test the difference between the z-transformed correlations for each model.

To examine changes in regional brain activation before and after CBT, group-level, between-session self-other comparisons were estimated using a linear mixed model with a subject-wise random intercept to account for subject-wise variance between pre- and post-treatment fMRI sessions in the clinical participants who completed treatment. Group-level, within-session GLMs were also performed to compare self-other estimates between patients and controls at pre- and post-treatment separately. Corrections for multiple comparisons were

performed using TFCE with 10000 permutations (Smith & Nichols, 2009; See Supplementary Materials). Reported clusters were thresholded at $p<0.001$ with a minimum size of 10 voxels.

Results

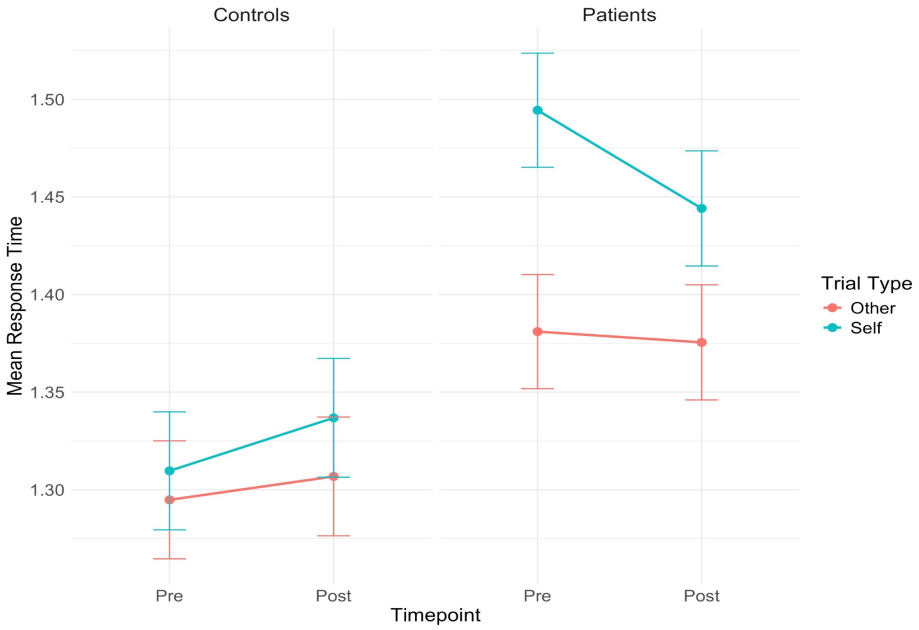
Behavioral Results

The mean percent symptom reduction after CBT across both patient groups was 53.70% ± 15.78 (Fang et al., 2024). Reaction time (RT) data were analyzed to determine changes in self vs. other processing across timepoints in patients and controls. The mean RTs and standard deviations for each group in pre- and post-treatment and for self- and other- trials are provided in Table S2. We conducted a linear mixed-effects model on the response times for self and other trials using fixed effects (time point, participant group, trial type) and random effects (participant intercept). This revealed a significant main effect of participant group, with patients showing higher response times compared to controls, $t(59.99)=4.39$, $p<.001$. There was also a significant main effect of time point, with post-treatment response times being higher than pre-treatment response times, $t(18260)=2.74$, $p<.01$. The main effect of trial type was not significant, $t(18230)=-1.60$, $p=.109$.

Significant interaction effects were also found between group and time point, $t(18270)=-5.50$, $p<.001$, and between group and trial type, $t(18230)=-7.55$, $p<.001$. The interaction between time point and trial type was not significant, $t(18230)=-1.11$, $p=.269$. However, the three-way interaction between group, time point, and trial type was significant, $t(18230)=3.09$, $p<.01$.

Figure 2

Three-Way Interaction of Time Point and Trial Type by Group.



Note. Mean response time (y-axis) for self trials in patients appears to be driving the interaction. Specifically, patients display a shorter response time for self- trials at post-treatment, and show a greater rate of change in self trials compared to other trials.

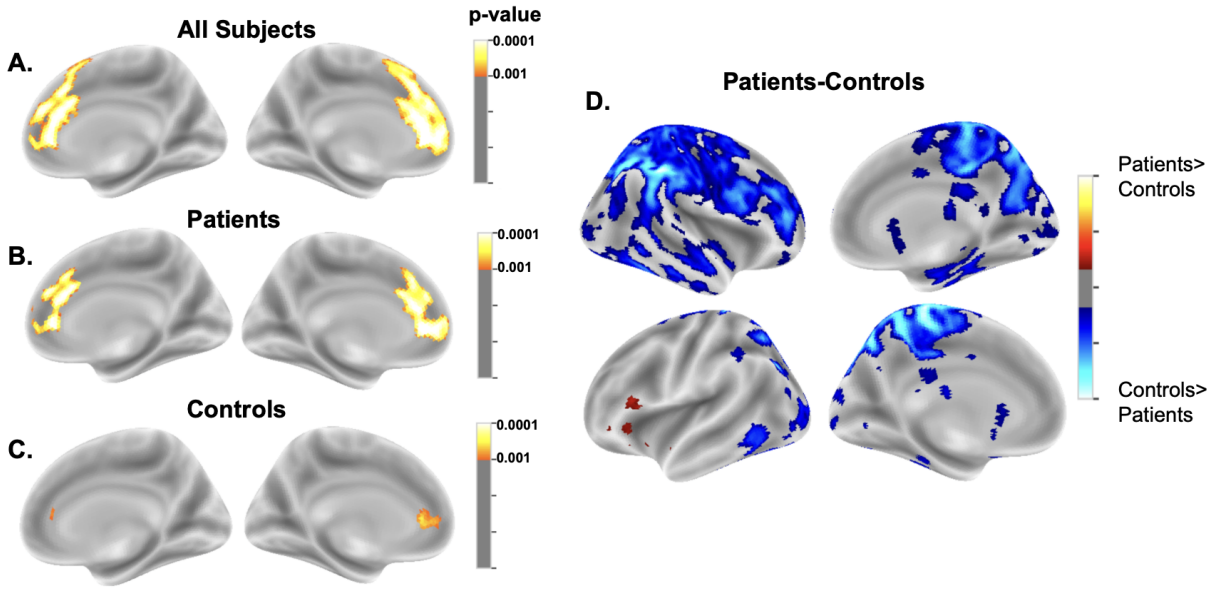
Neural Correlates of Self Versus Other Processing at Pre-Treatment

As our original hypotheses were specific to DN regions, we conducted ROI analyses using the DN mask to compare self-other processing within the mPFC, PCC, and bilateral TPJ regions of the DN within patients and controls across treatment (Table 1). At pre-treatment, there was a marginal group difference in self vs. other trials between patients and controls within the mPFC ($\beta = 0.71$, $p = 0.07$), which reflected increases in activity in self trials relative to fixation, and decreases in activity (greater deactivations) in other trials relative to fixation, in patients relative to controls.

Whole-brain analyses were also conducted to test differences in self vs. other trials across all subjects (Figure 3A) as well as separately in patients (Figure 3B) and healthy controls (Figure 3C). Across all subjects as well as within each group, the mPFC was consistently more activated during self when compared to other trials. The mPFC cluster in the self vs. other contrast was notably larger within patients compared to healthy controls (patient cluster size = 20621 mm^3; control cluster size = 4514 mm^3); however, this region did not survive corrections for multiple comparisons (Figure 3D).

Figure 3

Self Versus Other Whole-Brain Activations



Note. Whole-brain activations of the self versus other contrast at pre-treatment within A) all subjects, B) patients, and C) controls based on threshold-free cluster enhancement and thresholded at FWE-corrected $p<0.001$. Across all subjects as well as within each group, the mPFC displayed significantly higher activity during self compared to other trials. Uncorrected

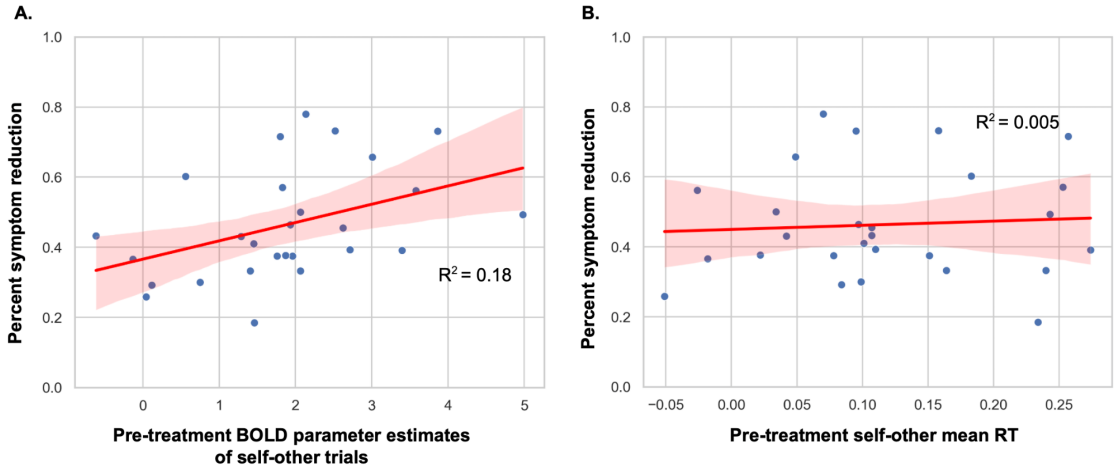
whole-brain differences between patients and controls can be seen in D); no regions survived thresholding at FWE-corrected $p<0.001$. Lateral views are not included for A-C due to no regions surviving correction thresholds within this view.

Self Versus Other Neural and Behavioral Measures Association with Treatment Response

For examining the relationship between the neural measures and treatment response, we extracted BOLD parameter estimates from whole-brain activity in the self vs. other contrast for all subjects at pre-treatment to use as predictors in a linear regression (see Supplemental Materials). We observed a significant association of the pre-treatment neural measures of self-other processing and treatment response (Figure 4A: $t = 2.34$, $p = 0.03$). Greater differences in BOLD responses for self vs. other trials at pre-treatment were associated with greater symptom reduction by post-treatment. In contrast, pre-treatment mean reaction times from self vs. other trials were not associated with treatment response (Figure 4B: $t = 0.34$, $p = 0.73$). Across groups, there was a main effect of time point on response time $t(59.99)=4.39$, $p<.001$. The neural model explained greater variance in treatment response ($R^2 = 0.18$) compared to the behavioral model ($R^2 = 0.005$). Fisher's Z test revealed a marginal difference between the two models ($z = 1.88$, $p = 0.06$).

Figure 4

Self Versus Other Pre-Treatment Neural and Behavioral Response



Note. A) Pre-treatment BOLD parameter estimates of self-other processing association with treatment response compared to B) behavioral self-other mean RT association with treatment response. Treatment response (y-axis) is measured as percent symptom reduction from pre- to post-treatment in patients. We found an overall greater associative strength in the neural correlate of self-other processing ($R^2 = 0.18$) when compared qualitatively to the behavioral model ($R^2 = 0.005$).

Changes in Self Versus Other Processing Between Pre- and Post-Treatment

ROI analyses from a linear mixed model examining pre-post changes in self vs. other trials in DN regions showed significant differences between patients and controls at post- versus pre-treatment in the left TPJ ($\beta = 0.91$, $p = 0.03$) and PCC ($\beta = 0.64$, $p = 0.02$), although neither region survived FDR corrections at $p<0.05$ for multiple comparisons.

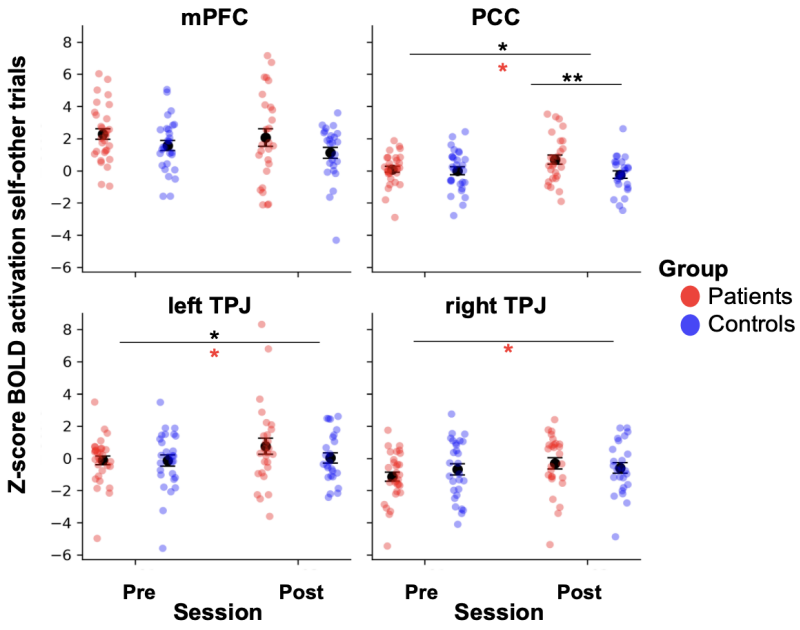
Despite null effects in the ROI analyses, we conducted several post-hoc tests to explore patterns of change in DN regions from pre-to-post in each group separately, and within each time point separately. Post-hoc tests revealed significant increases in activity in the self vs. other contrast in patients over the course of treatment in the left TPJ ($\beta = 0.94$, $p = 0.02$), right TPJ (β

= 0.78, p = 0.01), and PCC (β = 0.60, p = 0.02). As expected, we observed no significant differences in any DN regions within controls in the self vs. other contrast when comparing pre- and post-treatment (all p's > 0.43). At post-treatment, patients showed significantly greater activation in the PCC for self vs. other trials compared to controls (β = 0.96, p = 0.001), and marginal differences within both the mPFC (β = 0.95, p = 0.05) and left TPJ (β = 0.74, p = 0.07). These changes in activity did not correlate with symptom improvement from pre- to post-treatment (all p's > 0.26).

Whole brain analyses showed no significant differences between patients and controls at post- versus pre-treatment, which was contrary to our hypothesis. There were also no significant within group differences at post-treatment relative to pre-treatment in the whole brain analyses.

Figure 5

Exploratory Post-hoc Tests of Pre- and Post-Treatment



Note. Exploratory post-hoc tests of pre-post changes in self-other processing represented as z-score BOLD activation in self versus other trials in patients (red) and controls (blue) in the mPFC, PCC, left TPJ, and right TPJ. Black asterisks indicate significance in the between-group tests while red asterisks indicate significance in the within-group tests in the patient group (*: $p<0.05$, **: $p<0.001$). There were significant pre-post differences within patients in bilateral TPJ, and PCC (all p's < 0.02), and a significant group difference in the PCC at post-treatment ($p<0.001$), which survived $p_{FDR}<0.05$.

Table 1

Changes in Self Versus Other Processing from Pre- to Post-Treatment in DN Regions

Region	Pre-treatment			Post-treatment			Post – Pre-treatment		
	Patients Mean (SEM)	**Controls** Mean (SEM)	Patients–Controls β (p)	**Patients** Mean (SEM)	**Controls** Mean (SEM)	Patients–Controls β (p)	Post–Pre Patients β (p)	Post–Pre Controls β (p)	Patients–Controls, Post–Pre β (p)
mPFC	2.28(0.32)	1.57(0.32)	0.71(0.07)	2.06(0.54)	1.12(0.35)	0.95(0.05)	-0.25(0.71)	-0.67(0.93)	0.43(0.17)
PCC	0.09(0.18)	0.003(0.24)	0.07(0.38)	0.69(0.28)	-0.24(0.23)	**0.96(0.0006)***	**0.60(0.02)***	-0.26(0.81)	**0.64(0.02)**
lTPJ	-0.12(0.27)	-0.14(0.34)	0.00(0.5)	0.75(0.50)	0.02(0.31)	0.74(0.07)	**0.94(0.02)***	0.02(0.48)	**0.91(0.03)**
rTPJ	-1.15(0.28)	-0.69(0.35)	-0.50(0.94)	-0.31(0.35)	-0.60(0.33)	0.29(0.24)	**0.78(0.01)***	0.08(0.43)	0.29(0.23)

Note. Mean and standard error of the mean (SEM) values for patients and controls at pre- and post-treatment for each DN region of interest (mPFC, PCC, left TPJ, right TPJ). Significant differences ($p<0.05$) are bolded and asterisks indicate that the p-value survived FDR corrections at $p<0.05$.

Discussion

Our study investigated task-based neural correlates of SFA in patients with SAD, BDD, and healthy controls, and whether neural correlates of SFA would be stronger predictors of CBT response compared to behavioral correlates of SFA. We also examined changes in neural and behavioral measures of SFA during CBT. Although the SRET robustly activated DN regions in all participants, which replicated previous work using this task, we did not observe significant group differences in self-other processing at pre-treatment. We also did not observe significant changes in DN regions, or elsewhere, between pre- to post-treatment. However, pre-treatment activation in the mPFC during self-other processing was significantly associated with treatment response, whereas mean RT measures of self-other processing were not.

Our first hypothesis that patients would show greater activation of DN regions and have longer response latencies during self vs. other trials compared to controls was partially supported. Our results are consistent with DN areas being more activated in patients at pre-treatment, specifically in the mPFC, though these effects are only slight. These marginal findings are consistent with previous research (Dixon et al., 2022; Jamieson et al., 2023), except our sample was much more homogeneous with regard to SFA. Our results suggest that the absence of a group difference is likely unrelated to heterogeneity in SFA due to our SFA-focused sampling approach, but leaves open the possibility that other sources of heterogeneity in our transdiagnostic clinical sample could explain our pattern of findings, as well as the possibility that true differences in DN functioning, if they exist, are small. Due to our extensive scan time and number of trials, we do not expect these findings to exhibit Type II error. Rather, differences in DN activation may be more nuanced and perhaps better captured by examining the potential cumulative effects of SFA that accrue over time in the task. For instance, examining

perseverative effects of SFA (e.g., whether repeated endorsement of negative traits for self trials impacts subsequent trials as evidenced by ongoing DN activation in later trials) may be a more sensitive approach to examining the neural correlates of SFA.

With respect to our second hypothesis that neural correlates of SFA at pre-treatment would be more strongly associated with CBT response than behavioral RTs, we observed a significant association between pre-treatment neural correlates of SFA and subsequent treatment response. Pre-treatment neural activity within DN regions, such as the mPFC, may serve as a potential biomarker of CBT response and help identify candidates with SAD and BDD who are likely to respond to CBT. Better discrimination between self and other processing at pre-treatment (reflected in greater differences in BOLD responses to self vs. other trials) may be suggestive of one's capacity for taking on others' perspectives and having skills for modifying beliefs about others, which may support cognitive restructuring during CBT. Other studies showing differential brain activation during self vs. fix trials do not account for neural responses associated with thinking about anyone; thus, differences in brain activation during self vs. other capture neural processes that differ specifically between thinking about oneself vs. others. This hypothesis could be tested in future studies by linking self vs. other neural responses to measures of empathy. Another possibility is that greater pre-treatment activation in self vs. other processing reflects a mechanism that is malleable and supports treatment response in CBT. Repeated scanning of individuals throughout treatment may allow for further clarity on whether pre-treatment mPFC activation during self vs. other processing is a pre-requisite for successful CBT outcomes, a necessary mechanism of change during treatment, or both.

Additionally, our findings point to neural measures being potentially more meaningful than behavioral responses in predicting future clinical improvement from CBT. With regard to

prediction, several studies in anxiety and obsessive-compulsive disorders have focused predominantly on constructs such as emotion regulation and frontolimbic circuitry (Picó-Pérez et al., 2023), with relatively little explored in terms of non-threat-related circuitry. To our knowledge, this study is the first to highlight task-focused neural correlates related to self-referential processing in an anxiety-related sample, which may act independently or interact with threat mechanisms to predict treatment response.

The study's findings provide partial support for our third hypothesis regarding the relationship between changes in maladaptive SFA and DN activity during CBT. While we did not find robust differences in DN activation between patients and controls at pre-treatment, the observed changes in DN activity post-treatment within patient groups suggest a potential role for DN modulation in therapeutic response. This can be achieved using non-invasive neuromodulatory techniques, such as transcranial magnetic stimulation, transcranial direct current stimulation, or real-time neurofeedback. Our exploratory post-hoc analyses revealed sub-threshold differences in the changes of activity between patients and controls in the TPJ and PCC, though these changes were not significantly associated with clinical improvement. Surprisingly, these regions are distinct from those found in our first hypothesis (mPFC), and underscore the malleable nature of the DN across treatment.

Some limitations should be noted. First, the specific exclusion criteria required for the study may limit generalizability. Although our sampling approach was a strength of the study, it may also have unintentionally altered our results regarding DN activity. For instance, the DN activity being measured by relative increases or decreases from fixation may explain why most of our parameter estimates hovered around zero. In contrast, previous research with healthy controls (see Kelley et al. 2002) typically shows deactivations from baseline across all conditions

and simply less deactivation during self vs. other trials, which was not observed in this study. Furthermore, recent research has also highlighted the fractionation of the DN into multiple subsystems interacting regions of other networks (Andrews-Hanna et al., 2010). Therefore, our focus on examining the specific regions of the DN using this particular mask may have been too simplistic.

Despite these limitations, strengths include our transdiagnostic approach, which enhances the variability and relevance of SFA across different psychiatric disorders. By focusing on individuals across a spectrum of SAD and BDD symptomatology, our study provides a nuanced perspective on the neural underpinnings of maladaptive SFA.

Conclusions

Together, our results replicate previous work using similar tasks as the SRET that has shown a lack of difference in DN regions during self-referential processing between patient groups and healthy controls. Findings also support the notion of the neural correlates of SFA as a transdiagnostic biomarker of CBT response across SAD and BDD. The association between pre-treatment neural activity and treatment outcomes underscores the potential utility of targeting neural correlates of SFA in therapeutic interventions, regardless of specific diagnostic categories. Our study also highlights the complementary role of neural data in understanding SFA compared to behavioral measures alone. The significant association between pre-treatment neural correlates and treatment outcomes, compared to the lack of association with behavioral measures, suggests that neural markers may capture important aspects of SFA that behavioral measures cannot fully elucidate. Broadly, our findings contribute to the growing literature that neuroimaging-based biomarkers of CBT response may aid in treatment selection or treatment planning by identifying patients who are promising candidates for CBT. Future research should aim to replicate and

expand upon these findings in larger, more diverse samples, utilizing multimodal imaging approaches to comprehensively assess SFA across psychiatric disorders.

References

Abraham, A., Kaufmann, C., Redlich, R., Hermann, A., Stark, R., Stevens, S., & Hermann, C. (2013). Self-referential and anxiety-relevant information processing in subclinical social anxiety: An fMRI study. *Brain Imaging and Behavior*, *7*(1), 35–48. https://doi.org/10.1007/s11682-012-9188-x

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Allison, G. O., Kamath, R. A., Carrillo, V., Alqueza, K. L., Pagliaccio, D., Slavich, G. M., Shankman, S. A., & Auerbach, R. P. (2023). Self-referential Processing in Remitted Depression: An Event-Related Potential Study. *Biological Psychiatry Global Open Science*, *3*(1), 119–129. https://doi.org/10.1016/j.bpsgos.2021.12.005

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279. https://doi.org/10.1037/h0025907

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, *65*(4), 550–562. https://doi.org/10.1016/j.neuron.2010.02.005

Boehme, S., Miltner, W. H. R., & Straube, T. (2015). Neural correlates of self-focused attention in social anxiety. *Social Cognitive and Affective Neuroscience*, *10*(6), 856–862. https://doi.org/10.1093/scan/nsu128

Collins, A. C., & Winer, E. S. (2022). *Self-Referential Processing and Depression: A Systematic Review and Meta-Analysis*. https://doi.org/10.31219/osf.io/u7fw5

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Di Simplicio, M., Norbury, R., & Harmer, C. J. (2012). Short-term antidepressant administration reduces negative self-referential processing in the medial prefrontal cortex in subjects at risk for depression. *Molecular Psychiatry*, *17*(5), 503–510. https://doi.org/10.1038/mp.2011.16

Dixon, M. L., Moodie, C. A., Goldin, P. R., Farb, N., Heimberg, R. G., Zhang, J., & Gross, J. J. (2022). Frontoparietal and Default Mode Network Contributions to Self-Referential Processing in Social Anxiety Disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(1), 187–198. https://doi.org/10.3758/s13415-021-00933-6

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fang, A., Baran, B., Feusner, J. D., Phan, K. L., Beatty, C. C., Crane, J., Jacoby, R. J., Manoach, D. S., & Wilhelm, S. (2023). *Self-Focused Brain Predictors of Cognitive Behavioral Therapy Response in a Transdiagnostic Sample*. https://doi.org/10.1101/2023.08.30.23294878

First, M. B., Williams, J. B., Karg, R. S., & Spitzer, R. L. (2015). *SCID-5-RV*.

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Freton, M., Lemogne, C., Delaveau, P., Guionnet, S., Wright, E., Wiernik, E., Bertasi, E., & Fossati, P. (2014). The dark side of self-focus: Brain activity during self-focus in low and high brooders. *Social Cognitive and Affective Neuroscience*, *9*(11), 1808–1813. https://doi.org/10.1093/scan/nst178

Gaebler, M., Daniels, J., Lamke, J.-P., Fydrich, T., & Walter, H. (2014). Behavioural and neural correlates of self-focused emotion regulation in social anxiety disorder. *Journal of Psychiatry & Neuroscience*, *39*(4), 249–258. https://doi.org/10.1503/jpn.130080

Grimm, S., Ernst, J., Boesiger, P., Schuepbach, D., Hell, D., Boeker, H., & Northoff, G. (2009). Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. *Human Brain Mapping*, *30*(8), 2617–2627. https://doi.org/10.1002/hbm.20693

Hitchcock, P. F., Britton, W. B., Mehta, K. P., & Frank, M. J. (2023). Self-judgment dissected: A computational modeling analysis of self-referential processing and its relationship to trait mindfulness facets and depression symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(1), 171–189. https://doi.org/10.3758/s13415-022-01033-9

Hofmann, S. G. (2000). Self-focused attention before and after treatment of social phobia. *Behaviour Research and Therapy*, *38*(7), 717–725. https://doi.org/10.1016/S0005-7967(99)00105-9

Hofmann, S. G., Moscovitch, D. A., Kim, H.-J., & Taylor, A. N. (2004). Changes in Self-Perception During Treatment of Social Phobia. *Journal of Consulting and Clinical Psychology*, *72*(4), 588–596. https://doi.org/10.1037/0022-006X.72.4.588

Hofmann, S. G., & Otto, M. W. (2017). *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques* (2nd ed.). Routledge. https://doi.org/10.4324/9781315617039

Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin*, *107*(2), 156–176. https://doi.org/10.1037/0033-2909.107.2.156

Jamieson, A. J., Harrison, B. J., Delahoy, R., Schmaal, L., Felmingham, K. L., Phillips, L., & Davey, C. G. (2023). A brain model of altered self-appraisal in social anxiety disorder. *Translational Psychiatry*, *13*(1), 344. https://doi.org/10.1038/s41398-023-02644-7

Jones, N. P., Siegle, G. J., & Thase, M. E. (2008). Effects of Rumination and Initial Severity on Remission to Cognitive Therapy for Depression. *Cognitive Therapy and Research*, *32*(4), 591–604. https://doi.org/10.1007/s10608-008-9191-0

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the Self? An Event-Related fMRI Study. *Journal of Cognitive Neuroscience*, *14*(5), 785–794. https://doi.org/10.1162/08989290260138672

Komulainen, E., Heikkilä, R., Nummenmaa, L., Raij, T. T., Harmer, C. J., Isometsä, E., & Ekelund, J. (2018). Short-term escitalopram treatment normalizes aberrant self-referential processing in major depressive disorder. *Journal of Affective Disorders*, *236*, 222–229. https://doi.org/10.1016/j.jad.2018.04.096

LeMoult, J., Kircanski, K., Prasad, G., & Gotlib, I. H. (2017). Negative Self-Referential Processing Predicts the Recurrence of Major Depressive Episodes. *Clinical Psychological Science*, *5*(1), 174–181. https://doi.org/10.1177/2167702616654898

Liebowitz, M. R. (2013). *Liebowitz Social Anxiety Scale* [Dataset]. https://doi.org/10.1037/t07671-000

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, *128*(4), 638–662. https://doi.org/10.1037/0033-2909.128.4.638

Phillips, K. A., Hollander, E., Rasmussen, S. A., Aronowitz, B. R., DeCaria, C., & Goodman, W. K. (1997). A severity rating scale for body dysmorphic disorder: Development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. *Psychopharmacology Bulletin*, *33*(1), 17–22.

Picó-Pérez, M., Fullana, M. A., Albajes-Eizagirre, A., Vega, D., Marco-Pallarés, J., Vilar, A., Chamorro, J., Felmingham, K. L., Harrison, B. J., Radua, J., & Soriano-Mas, C. (2023). Neural predictors of cognitive-behavior therapy outcome in anxiety-related disorders: A meta-analysis of task-based fMRI studies. *Psychological Medicine*, *53*(8), 3387–3395. https://doi.org/10.1017/S0033291721005444

Reilly, E. E., Lavender, J. M., Berner, L. A., Brown, T. A., Wierenga, C. E., & Kaye, W. H. (2019). Could repetitive negative thinking interfere with corrective learning? The example of anorexia nervosa. *International Journal of Eating Disorders*, *52*(1), 36–41. https://doi.org/10.1002/eat.22997

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact

in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Scheier, M. F., & Carver, C. S. (1985). The Self-Consciousness Scale: A Revised Version for Use with General Populations [1]. *Journal of Applied Social Psychology*, *15*(8), 687–699. https://doi.org/10.1111/j.1559-1816.1985.tb02268.x

Schneider, R. L., Arch, J. J., & Wolitzky-Taylor, K. B. (2015). The state of personalized treatment for anxiety disorders: A systematic review of treatment moderators. *Clinical Psychology Review*, *38*, 39–54. https://doi.org/10.1016/j.cpr.2015.02.004

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Simpson, H. B., Maher, M., Page, J. R., Gibbons, C. J., Franklin, M. E., & Foa, E. B. (2010). Development of a Patient Adherence Scale for Exposure and Response Prevention Therapy. *Behavior Therapy*, *41*(1), 30–37. https://doi.org/10.1016/j.beth.2008.12.002

Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Thesen, S., Heid, O., Mueller, E., & Schad, L. R. (2000). Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine*, *44*(3), 457–465. https://doi.org/10.1002/1522-2594(200009)44:3<457::AID-MRM17>3.0.CO;2-R

Tisdall, M. D., Hess, A. T., Reuter, M., Meintjes, E. M., Fischl, B., & Van Der Kouwe, A. J. W. (2012). Volumetric navigators for prospective motion correction and selective

reacquisition in neuroanatomical MRI. *Magnetic Resonance in Medicine*, *68*(2), 389–399. https://doi.org/10.1002/mrm.23228

Wang, S., Tepfer, L. J., Taren, A. A., & Smith, D. V. (2020). Functional parcellation of the default mode network: A large-scale meta-analysis. *Scientific Reports*, *10*(1), 16096. https://doi.org/10.1038/s41598-020-72317-8

Wilhelm, S., Phillips, K. A., Didie, E., Buhlmann, U., Greenberg, J. L., Fama, J. M., Keshaviah, A., & Steketee, G. (2014). Modular Cognitive-Behavioral Therapy for Body Dysmorphic Disorder: A Randomized Controlled Trial. *Behavior Therapy*, *45*(3), 314–327. https://doi.org/10.1016/j.beth.2013.12.007

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635