A Brain Model of Disturbed Self-Appraisal in Social Anxiety Disorder

*Alec J. Jamieson[1], Ben J. Harrison[1], Rebekah Delahoy[1], Lianne Schmaal[2,3], Kim L. Felmingham[4], Lisa Phillips[4], *Christopher G. Davey[5]

[1]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Parkville, Australia

[2]Orygen, Parkville, Australia

[3]Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[4]Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia

[5]Department of Psychiatry, The University of Melbourne, Parkville, Australia

***Corresponding authors**: Dr. Alec Jamieson. Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne. Level 3, 161 Barry St. Carlton, VIC 3053.

Email: alec.jamieson@unimelb.edu.au

Professor Christopher Davey. Department of Psychiatry, The University of Melbourne. Level 1 North Block, Royal Melbourne Hospital, Parkville, VIC 3052 **T:** +61 3 8344 5509 **E:** c.davey@unimelb.edu.au

Word count: 3995

Abstract

The brain's default mode network has a central role in the processing of information concerning oneself. Dysfunction in this self-referential processing represents a key component of multiple mental health conditions, including social anxiety disorder (SAD). This study aimed to clarify alterations to network dynamics present during self-appraisal in SAD participants. Thirty-eight adolescents and young adults with SAD and 72 healthy control participants underwent a self-referential processing fMRI task. The task involved two primary conditions of interest: *direct self-appraisal* (thinking about oneself) and *reflected self-appraisal* (thinking about how others might think about oneself). Dynamic causal modelling and parametric empirical Bayes were then used to explore differences in the effective connectivity of the default mode network between groups. We observed connectivity differences between SAD and healthy control participants in the reflected self-appraisal but not the direct self-appraisal condition. Specifically, SAD participants exhibited greater excitatory connectivity from the posterior cingulate cortex (PCC) to medial prefrontal cortex (MPFC) and greater inhibitory connectivity from the inferior parietal lobule (IPL) to MPFC. In contrast, in the absence of task modulation, SAD participants exhibited reduced intrinsic connectivity, with reduced excitatory connectivity from the PCC to MPFC and reduced inhibitory connectivity from the IPL to MPFC. Participants with SAD showed changes to MPFC inputs that occurred during reflected self-appraisal that were also present intrinsically. The presence of connectivity differences in reflected and not direct self-appraisal is consistent with the characteristic fear of negative social evaluation that is experienced by people with SAD.

Introduction

Social anxiety disorder (SAD) is one of the most common of all mental illnesses, with a lifetime prevalence of 11% (1). It has a median age of onset in early adolescence (2, 3) and peak prevalence later in adolescence (4). People with SAD often experience significant distress during social encounters (5), and in turn, make considerable efforts to avoid them. As a result of these factors, SAD has a dramatic effect on social and educational functioning throughout a critical period of development (6) resulting in reduced quality of life (7).

An essential characteristic of SAD is a consistent fear of negative evaluation, which can result in excessively focusing on themselves during social interactions. Their self-representations are influenced by how they imagine their appearance and behavior are perceived by others, with fears that they are viewed as socially inept, gauche, and awkward (8). This disturbance in self-appraisal processes is associated with marked anxiety, and is one of the features of SAD that is targeted by cognitive behavioral therapy (9).

Functional MRI studies of participants with SAD often show alterations in brain regions that support self-appraisal processes. The medial prefrontal cortex (MPFC) and posterior cingulate cortex (PCC) are core regions of the default mode network (DMN), a collection of brain regions which are highly active during self-appraisal (10, 11). SAD associated alterations have been demonstrated in both regions, with the MPFC showing increased activation during self-appraisal processes (12, 13) and the PCC showing increased activation in response to facial stimuli (14, 15). Similarly, altered functional connectivity between the MPFC and PCC has been illustrated in SAD participants (16, 17), although the directionality of these results is inconsistent between studies.

In addition to these midline regions, self-appraisal is associated with increased activity and connectivity with a third component of the DMN, the inferior parietal lobule (IPL). Using effective connectivity (the directional effect that activity in one region has on the activity of another), we have previously shown that the PCC exerts a positive influence on the IPL and MPFC during self-

appraisal, with the latter in turn exerting negative feedback on the PCC (11). In a more recent study, and using the same modelling approach, our group compared direct self-appraisal (where the participant assessed whether personality adjectives described them) with reflected self-appraisal (where the participant assessed whether *other* people would think personality adjectives described them) (18). The latter task component showed similar patterns of directed connectivity, although with greater activation of the PCC and IPL, and a distinct pattern of negative influence of the IPL on PCC.

The examination of reflected self-appraisal is likely to be particularly pertinent for people with SAD, in light of the marked anxiety associated with their fear of negative appraisal by others that is inherent to the condition. Our aim was to examine brain connectivity in participants with SAD as they engaged in a task that examined directed and reflected self-appraisal processes. As in our previous studies, we aimed to use dynamic causal modelling to examine effective connectivity, comparing participants with SAD with control participants across the two conditions. We hypothesized that connectivity differences would be more apparent for SAD participants performing reflected self-appraisal than direct self-appraisal. Consistent with our previous work (19), we propose that these differences will likely occur in the modulation from the MPFC to PCC.

Methods and Materials

Participants

Forty-two unmedicated SAD participants were recruited from outpatient mental health clinics in the western and northern suburbs of Melbourne, Australia. These participants had a primary diagnosis of SAD, as assessed with the Structured Clinical Interview for DSM-5 Axis I disorders (SCID-5-RV; 20). SAD participants' symptoms were in the moderate to severe range on the Liebowitz Social Anxiety Scale (LSAS; 21). Comorbid MDD was not an exclusion criterion as long as the severity of depressive symptoms was mild-to-moderate, quantified by a Montgomery-Åsberg Depression

Rating Scale score ≤ 20 (22). Imaging data from 4 SAD participants were excluded due to excessive head motion (1), and suboptimal performance of the external attention component of the task (3; see below), leaving a total sample of 38 SAD participants.

We recruited 122 healthy controls with no current or past diagnoses of mental illness (assessed with the SCID-5 non-patient interview; First et al., 2015), which we have reported on previously (18). Exclusion due to excessive head motion (N = 3), inadequate task performance (N = 8), and insufficient subject level activation for DCM (N = 7), resulted in 104 participants with adequate data. Due to differences in age between the healthy control and SAD samples, we used propensity score matching to select a subsample of 72 controls who were matched with the SAD participants on age and gender (see Supplementary Materials for details).

All participants met the following inclusion criteria: they were aged between 16 and 25 years, were competent English speakers, had no current treatment with psychoactive medications, had no dependence on alcohol or other drugs, showed no incidental neurological findings on MRI, and had no further contraindications to MRI. Written informed consent was obtained from each participant following a complete description of the study protocol. For participants under 18 years, both participant assent and parent consent were obtained. The study protocol was approved by The University of Melbourne Human Research Ethics Committee and was carried out in accordance with the Declaration of Helsinki.

Task design

Participants completed an fMRI task composed of 4 experimental conditions: direct self-appraisal, reflected self-appraisal, external attention, and rest-fixation – the same task design described by Delahoy and colleagues (18). In 3 of the conditions, participants were presented with words drawn from a list of frequently used trait adjectives (23). To heightening self-reflection, words were selected to be neither extremely favourable nor unfavourable. As such, selected adjectives were from the subset of words rated as most ‘meaningful’ and were distributed around the median rating

for ‘likeableness’. Words were presented in blocks (6 blocks per condition, 5 words per block) and the word order was randomized for each participant. In the direct self-appraisal condition, they were asked to respond to the question “Would you use this word to describe you?”. In the reflected self-appraisal condition, they were asked “Would others use this word to describe you?”, and in the external attention condition they responded to the question “Does this word have four or more vowels?”. To each stimulus, they responded “Yes” or “No” by pressing one of two allocated buttons on a curved 4-button fiber-optic response pad (Cambridge Research Systems Ltd.). The 3 lists of 30 words that formed these conditions were matched on valence and number of vowels and were counterbalanced across participants (Supplementary Table S1). Each 27-second block (2 seconds of instruction followed by 5 words presented for 5 seconds each) was interspersed with a 10-second rest-fixation block in which participants were asked to fixate on a centrally presented crosshair.

Image acquisition

A 3-T General Electric Discovery MR750 system equipped with an 8-channel phased-array head coil was used in combination with ASSET parallel imaging. The functional sequence consisted of a single-shot gradient-recalled echo planar imaging sequence in the steady state (repetition time, 2 s; echo time, 33 ms; pulse angle, 90º) in a 23 cm field of view, with a 64x64 pixel matrix and a slice thickness of 3.5 mm (no gap). Thirty-six interleaved slices were acquired parallel to the anterior-posterior commissure line with a 20º anterior tilt to better cover ventral prefrontal cortical regions. The total sequence time was 11 min 24 s, corresponding to 342 whole-brain echo-planar imaging volumes. A T1-weighted high-resolution anatomical image was acquired for each participant to assist with functional time-series co-registration (140 contiguous slices; repetition time, 8.2 ms; echo time, 3.2 ms; flip angle, 13º; in a 25.6 cm field-of-view, with a 256x256 pixel matrix and a slice thickness of 1 mm). To assist with noise reduction and head immobility, participants were fitted with insert-ear protection and their heads were supported with foam-padding inserts.

Image preprocessing

Imaging data were transferred to a Unix-based platform that ran MATLAB version 9.3 (The MathWorks Inc., Natick, USA) and Statistical Parametric Mapping Version 12 v7771 (SPM12; Wellcome Trust Centre for Neuroimaging, UK). Each participant's time-series were aligned to the first image using least-squares minimization and a 6-parameter rigid-body spatial transformation. Motion Fingerprint toolbox (Wilke, 2012) was used to characterize scan-to-scan head motion for each participant. Participants' data were excluded if scan-to-scan displacement exceeded 2.5 mm (~1 native voxel) or total displacement exceeded 3 mm. Realigned functional images were then co-registered to each participant's respective T1 anatomical scans, which were segmented and spatially normalized to the International Consortium of Brain Mapping template using the unified segmentation approach. The functional images were transformed to 2 mm isotropic resolution and were smoothed with an 8 mm full-width-at-half-maximum Gaussian filter.

General linear modelling

Following preprocessing, a general linear model (GLM) approach in SPM12 was used to identify regions that were significantly activated by the different conditions of the task. This was done by specifying the durations and onsets of each block for the direct self-appraisal, reflected self-appraisal, rest-fixation, and external attention conditions, which were then convolved with a canonical hemodynamic response function. The external attention condition was considered an appropriate baseline as it was matched with the self-appraisal conditions on stimulus features yet required specific attentional demands to minimize the intrusion of task-independent thought. A high-pass filter (1/128 s) accounted for low-frequency drifts. GLM with local autocorrelation correction was used to calculate parameter estimates at each voxel. Primary contrast images were carried forward to second-level analyses using the summary statistics approach to random-effects analyses. Independent sample *t*-tests were used for between-groups analyses, with a whole brain threshold family-wise error rate (FWE) corrected threshold of $P < 0.05$, $K_E \geq 20$ voxels.

Following Delahoy and colleagues (18), we then applied conjunction (null) analyses to identify core-self DMN regions across groups: those that were commonly activated by both self-appraisal conditions as well as rest-fixation, but that showed further distinct activation during the self-appraisal conditions relative to rest-fixation. These core-self regions were used as volumes of interest (VOIs) for further analysis with DCM. Results are reported in Montreal Neurological Institute space.

Dynamic causal modelling

DCM is a method for estimating the directional interactions between brain regions from observed neuroimaging data (24). For tasks, this estimation is demarcated into invariant connectivity in the absence of task modulation (intrinsic connectivity) and connectivity that is modulated by experimental stimuli (25). For interactions between regions, the strength of this relationship is measured in hertz (Hz) and positive values indicate excitation while negative values indicate inhibition. In contrast, self-connections are unitless log scaling parameters for which positive values indicate greater inhibition whereas negative values indicate reduced inhibition.

As in our previous studies (11, 18), we used a conjunction analysis to identify peak coordinates for what we have termed “core-self DMN regions”. These are regions that are active during rest-fixation and self-appraisal, and include the MPFC, PCC and IPL. The time-series for our chosen VOIs were then extracted at a single subject level using the first eigenvariate of voxels within 5 mm of the subject-specific maxima that showed significant activation for the rest-fixation and direct and reflected self-appraisal > external attention contrast ($P < 0.05$). If a given VOI had inadequate activation, the threshold was incrementally relaxed up to $P < 0.5$ (25). Specific coordinates for all three regions were personalized to each subject’s specific local maximum, which were required to occur within 8 mm of the group level peaks (see Supplementary Figure S1). A full model was specified with intrinsic bidirectional connections between the MPFC, PCC, and IPL (Supplementary Figure S2) and modulation for both the direct and reflected self-appraisal

conditions occurring for all connections. Direct input into the network was modeled using the overall 'broad self' (direct + reflected + rest) into PCC and the input matrix was not mean centered. This full model was then estimated separately for each subject.

We deployed Parametric Empirical Bayes (PEB) to examine the differences in connectivity parameters between healthy controls and SAD participants (26). PEB enables the inclusion of the estimated variance of connectivity parameters when investigating between-group effects, thereby allowing for more reliable parameter estimates than using classical tests. While Bayesian statistics has no concept of statistical significancy, thresholding can be used to identify effects which are likely of interest. We used a posterior probability (PP) threshold of greater than 0.95 (27). The regressors included in this analysis represented 1) the average connectivity across all participants, 2) the effect of SAD, and 3) the effect of age. An automatic search over nested PEB models was conducted following the estimation of a group level PEB model (28, 29). After the final iteration, Bayesian model averaging was performed to determine the strength of connections in the last Occam's window of 256 models.

Leave-one-out cross-validation

To assess the predictive validity of parameters which demonstrated between group differences we used leave-one-out cross-validation (LOOCV) across both groups. In the PEB framework, LOOCV aims to determine whether the size of between-group effects on parameters is sufficiently large to predict a variable of interest (27), in this case, anxiety symptom severity. This does so by estimating a group-level PEB model while excluding one subject, then using this PEB model to predict the left-out subject's anxiety symptom severity. This predicted anxiety score is then correlated with the observed anxiety symptom severity. A significant correlation between the expected and observed values demonstrates that the effect size was sufficiently large to predict the left-out subjects' anxiety symptom severity above chance (see (27) for further details). Anxiety symptom severity was

assessed through the LSAS (social anxiety symptoms), as well as the State-Trait Anxiety Inventory (STAI; 30) General and Present Subscales (trait and state anxiety symptoms respectively).

Results

Clinical and demographic characteristics

Comparison of clinical characteristics revealed that SAD participants had significantly higher LSAS ($t(56.26) = -13.65$, $P < .001$), STAI General ($t(108) = -13.52$, $P < .001$) and STAI Present scores ($t(108) = -9.49$, $P < .001$) compared with healthy controls (Table 1). For distribution of anxiety severity scores across groups see Supplementary Figure S3. No significant differences were observed between groups for age, gender, or IQ (Table 1).

Task performance

There was a statistically significant difference in the RT between the three condition types ($F(1.28\ 138.43) = 34.01$, $P < .001$) across groups. Post hoc testing revealed significantly slower RTs during the external attention condition compared with both the direct ($P < .001$) and reflected self-appraisal conditions ($P < .001$), as well as a small but significantly slower RT for reflected compared with direct self-appraisal ($P = .019$). There was no significant main effect of diagnosis on RT ($F(1,108) = 1.21$, $P = .274$), nor a significant interaction effect between diagnostic status and RT ($F(1.28\ 138.43) = 1.64$, $P = .204$). Mean accuracy for the external attention task was 94.4% for controls and 94.0% for SAD participants, with no significant difference between groups ($P = .868$; Table 1). SAD participants were also more likely to answer “Yes” to the question “Does this word describe you?” ($t(108) = -2.68$, $P = 0.008$). While there were no significant differences between groups in the percentage of “Yes” responses to the question “Would others use this word to describe you?”, it should be noted that this effect was on the boundary of the statistical significance threshold ($t(108) = -1.98$, $P = 0.05$).

Mapping common and distinct activations across self-appraisal and rest

Across both healthy controls and SAD participants, direct and reflected self-appraisal conditions produced significant and broadly similar activation patterns which encompassed core regions of the DMN (Figure 1A/B and Figure 2A/B, respectively): namely the anterior medial wall cortex (including frontopolar MPFC), posteromedial cortex (incorporating the PCC and precuneus), and IPL.

The three-way conjunction analysis across participants to identify core-self regions revealed three significant regional clusters: MPFC (peak coordinate, x = -10, y = 54, z = 10; cluster size = 630; peak *t*- value = 7.15), PCC (peak coordinate, x = -6, y = -50, z = 26; cluster size = 158; peak *t*-value = 7.24), and left IPL (peak coordinate, x = -48, y = -60, z = 24; cluster size = 184; peak *t*-value = 6.63; Supplementary Figure S1).

Overall, SAD participants and healthy controls demonstrated similar activation patterns during the direct and reflected self-appraisal conditions. At a whole-brain level, between-groups comparisons revealed no significant differences in regional activation for the main contrasts of interest ($P_{FWE} < 0.05$). At a more lenient threshold ($P < 0.001$, uncorrected), differences were observed for both the direct self-appraisal > external attention and reflected self-appraisal > external attention contrasts. During direct self-appraisal, SAD participants demonstrated decreased activation in the left pars triangularis of the inferior frontal gyrus, left premotor cortex, and ventral frontal eye fields compared with healthy control participants (Figure 1C; Supplementary Table S2). During reflected self-appraisal SAD participants illustrated decreased activity in the right pars opercularis of the inferior frontal gyrus, left pars triangularis of the inferior frontal gyrus, and left premotor cortex (Figure 2C; Supplementary Table S2).

Connectivity differences between SAD participants and healthy controls

SAD participants had reduced intrinsic excitatory connectivity from the PCC to MPFC (expected value = -.09 Hz, PP = 1.00) and reduced intrinsic inhibitory connectivity from the IPL to MPFC (expected value = .08 Hz, PP = 1.00) in comparison to healthy control participants (Figure 3A).

During reflected self-appraisal, SAD participants had greater excitatory connectivity from the PCC to MPFC (expected value = .05 Hz, PP = .95) and greater inhibition from the IPL to MPFC (expected value = -.06 Hz, PP = .97) in comparison with healthy controls participants (Figure 3C). No such differences between groups were observed for direct self-appraisal modulation. The expected values and PP for all parameters estimates are reported in Supplementary Table S3.

Leave-one-out cross-validation

Using those parameters which demonstrated between groups differences we performed a LOOCV analysis in the PEB framework to determine whether the size of this effect on these parameters could significantly predict anxiety symptom severity. The intrinsic connectivity from the IPL to MPFC and from the PCC to MPFC resulted in a significant out-of-samples correlation between the predicted and observed STAI general scores (r = .25, P = .004; Figure 4A). No significant correlations were found for these intrinsic parameters and LSAS or STAI present scores (Figure 4B and C). For the full list of individual parameter correlations see Supplementary Table S4.

Discussion

The current study investigated a brain model of the network dynamics present during self-referential processing in young people with SAD. We observed DMN connectivity differences in SAD participants compared with healthy controls during reflected, but not direct self-appraisal. Specifically, SAD participants illustrated greater excitatory connectivity from the PCC to MPFC and greater inhibitory connectivity from the IPL to MPFC. The specificity of this finding for the

reflected self-appraisal condition is consistent with the nature of SAD, whereby sufferers show acute, anxiety-inducing preoccupation with the views that others hold of them.

The PCC has been highlighted as a key region in coordinating the function of the DMN during social processes (31). Our previous work hypothesized that representations of the narrative self, the aspect underlying direct and reflected self-appraisal, are primarily generated by the PCC (32). In contrast to direct self-appraisal, however, reflected self-appraisal appears to be marked by greater regulatory feedback from the IPL due to its greater dependence on perspective taking and episodic memory retrieval (18). The role of the MPFC has been suggested to gate these representations of the self into conscious awareness (11, 33) and to be more broadly involved in evaluative processes (34, 35).

Cognitive models of social anxiety have proposed that increased self-focused attention is a core feature in the development and maintenance of this disorder (36). A heightened level of self-focused attention has been shown to correlate with the activity of the MPFC and PCC in participants who are highly socially anxious (37). During reflected self-appraisal, the increased modulation from the IPL and PCC to MPFC might indicate a heightened focus on how the self is perceived to be represented by others in those with SAD. Interestingly, in SAD participants the connectivity from the PCC to MPFC during reflected self-appraisal also appears more similar to the modulation present during direct self-appraisal. This supports the idea that for SAD participants reflected self-appraisal is more dependent on their constructed narrative self, in contrast to how this process functions in healthy participants (18). CBT has been shown to reduce self-focused negative thoughts and beliefs (9), with decreased frequency of negative thinking being associated with reduced social anxiety (38). As such, examining whether these changes in connectivity predict response to CBT may be of interest for future research.

While SAD participants have demonstrated both hyper- and hypoconnectivity of the DMN across studies, more recent work has shown reduced resting-state connectivity within the DMN (39, 40).

Reduced DMN connectivity has also been associated with increased trait anxiety scores (41). Broadly, these observations and our own are consistent with the Topography of the Anxious Self model proposed by Angeletti and colleagues (42). In short, this model suggests that abnormal resting-state connectivity within the DMN and between the DMN and salience network may contribute to trait features of anxiety; these baseline alterations, in turn, increase the sensitivity to task associated hyperactivity of these same regions and increases the likelihood that specific stimuli induce state anxiety. Moreover, it predicts that as there are greater reductions to within-DMN connectivity, these regions will exhibit stronger response to symptom-sensitive stimuli. Broadly, these observations and our own are consistent with the Topography of the Anxious Self model proposed by Angeletti and colleagues (42). This interpretation is additionally supported by the predictive validity of the intrinsic parameter for trait anxiety symptoms.

Notably we did not observe increased activation of the MPFC, which has been reported in previous investigations of self-referential processing in anxiety disorders (12, 13). One potential explanation for these differences was that our sample of SAD participants was relatively young, and as a result has not experienced the level of chronicity reported in previous work. Prolonged exposure to stress and anxiety has been shown to result in changes to the structure and function of the prefrontal cortex (43). As such, altered activity of the MPFC may be a product of prolonged exposure to high levels of anxiety rather than the underlying cause of these symptoms. Moreover, adolescence is a key developmental period in the maturation of the MPFC (44, 45), with previous work identifying reduced MPFC activity during self-processing in adults compared to adolescents (46). Longitudinal examination of these effects will be necessary to clarify whether this heighten baseline activity may be masking MPFC hyperactivity.

This study uses novel methodology to investigate altered directional interactions of the DMN in SAD, however, it is not without its limitations. The block design of the task and the selection of trait adjectives distributed around the median rating for ‘likeableness’ meant that we were unable to disentangle if the observed effects may have been different for positive and negatively valence

adjectives. While recent work has illustrated no significant differences in DMN activity between positive and negative direct self-referential processing (47), there may be differences present during reflected self-referential processing. The SAD sample used in this study was also relatively small, which makes replication in a larger sample of adolescents and young adults with SAD critical.

Conclusions

Our study has provided a novel brain model of DMN dysfunction present in SAD participants during self-referential processing. We observed SAD associated reductions in the intrinsic connectivity to the MPFC as well as greater modulation to the MPFC specifically during reflected self-referential processing. These observations highlight anxiety associated alterations to connectivity which occur during reflected self-appraisal. We also demonstrated that effect size of these alterations in the intrinsic connectivity were associated with trait anxiety symptoms. Further investigations into how psychological treatments which specifically target this appraisal process, such as CBT, alters this dysfunction may provide key insight into the neurobiology of SAD.

CRediT Authorship Contribution Statement

AJJ: Conceptualization; Formal analysis; Methodology; Writing - original draft; Visualization. BJH: Conceptualization, Funding acquisition; Methodology, Supervision; Writing - review & editing. RD: Formal analysis; Writing - review & editing. LS: Resources; Writing - review & editing. KLF: Resources; Writing - review & editing. LP: Resources; Writing - review & editing. CGD: Conceptualization; Funding acquisition; Methodology; Supervision; Writing - review & editing.

Declaration of interest

None of the authors have competing interests to declare.

Funding

The study was supported by National Health and Medical Research Council (NHMRC) Project Grants 1064643 and 1145010 (principal investigator, BJH). CGD and BJH were supported by NHMRC Career Development Fellowships (1141738 and 1124472, respectively).

Acknowledgments

The authors thank Dr. Laura Finlayson-Short, Dr. Yara Toenders, Dr. Hannah Savage and Ms. Lisa Incerti for their contributions to data collection. We also thank the staff from the Sunshine Hospital Medical Imaging Department (Western Health, Melbourne).

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen HU (2012): Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int J Meth Psych Res*. 21:169-184.

2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005): Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 62:593-602.

3. McEvoy PM, Grove R, Slade T (2011): Epidemiology of anxiety disorders in the Australian general population: findings of the 2007 Australian National Survey of Mental Health and Wellbeing. *Aust N Z J Psychiatry*. 45:957-967.

4. Costello EJ, Copeland W, Angold A (2011): Trends in Psychopathology across the Adolescent Years: What Changes When Children become Adolescents, and When Adolescents become Adults? *J Child Psychol Psyc*. 52:1015-1025.

5. Mesa F, Beidel DC, Bunnell BE (2014): An Examination of Psychopathology and Daily Impairment in Adolescents with Social Anxiety Disorder. *Plos One*. 9.

6. Leigh E, Clark DM (2018): Understanding Social Anxiety Disorder in Adolescents and Improving Treatment Outcomes: Applying the Cognitive Model of Clark and Wells (1995). *Clin Child Fam Psychol Rev*. 21:388-414.

7. Stein MB, Stein DJ (2008): Social anxiety disorder. *Lancet*. 371:1115-1125.

8. Rapee RM, Heimberg RG (1997): A cognitive-behavioral model of anxiety in social phobia. *Behav Res Ther*. 35:741-756.

9. Gregory B, Peters L (2017): Changes in the self during cognitive behavioural therapy for social anxiety disorder: A systematic review. *Clin Psychol Rev*. 52:1-18.

10. Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006): Self-referential processing in our brain--a meta-analysis of imaging studies on the self. *Neuroimage*. 31:440-457.

11. Davey CG, Pujol J, Harrison BJ (2016): Mapping the self in the brain's default mode network. *Neuroimage*. 132:390-397.

12. Blair K, Geraci M, Devido J, McCaffrey D, Chen G, Vythilingam M, et al. (2008): Neural response to self- and other referential praise and criticism in generalized social phobia. *Arch Gen Psychiatry*. 65:1176-1184.

13. Blair K, Geraci M, Otero M, Majestic C, Odenheimer S, Jacobs M, et al. (2011): Atypical modulation of medial prefrontal cortex to self-referential comments in generalized social phobia. *Psychiatry Res*. 193:38-45.

14. Choi SH, Shin JE, Ku J, Kim JJ (2016): Looking at the self in front of others: Neural correlates of attentional bias in social anxiety. *Journal of Psychiatric Research*. 75:31-40.

15. Schneier FR, Rodebaugh TL, Blanco C, Lewin H, Liebowitz MR (2011): Fear and avoidance of eye contact in social anxiety disorder. *Compr Psychiat*. 52:81-87.

16. Bruhl AB, Delsignore A, Komossa K, Weidt S (2014): Neuroimaging in social anxiety disorder-A meta-analytic review resulting in a new neurofunctional model. *Neurosci Biobehav R*. 47:260-280.

17. Mizzi S, Pedersen M, Lorenzetti V, Heinrichs M, Labuschagne I (2022): Resting-state neuroimaging in social anxiety disorder: a systematic review. *Molecular Psychiatry*. 27:164-179.

18. Delahoy R, Davey CG, Jamieson AJ, Finlayson-Short L, Savage HS, Steward T, et al. (2022): Modulation of the brain's core-self network by self-appraisal processes. *Neuroimage*. 251:118980.

19. Davey CG, Breakspear M, Pujol J, Harrison BJ (2017): A Brain Model of Disturbed Self-Appraisal in Depression. *Am J Psychiat*. 174:895-903.

20. First MB, Williams JB, Karg RS, Spitzer RL (2015): Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV). *Arlington, VA: American Psychiatric Association*.

21. Liebowitz MR (1987): Social phobia. *New York, NY: Guilford Publications*.

22. Montgomery SA, Asberg M (1979): A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 134:382-389.

23. Anderson NH (1968): Likableness ratings of 555 personality-trait words. *Journal of personality and social psychology*. 9:272.

24. Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. *Neuroimage*. 19:1273-1302.

25. Zeidman P, Jafarian A, Corbin N, Seghier ML, Razi A, Price CJ, et al. (2019): A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *Neuroimage*. 200:174-190.

26. Friston KJ, Litvak V, Oswal A, Razi A, Stephan KE, van Wijk BCM, et al. (2016): Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage*. 128:413-431.

27. Zeidman P, Jafarian A, Seghier ML, Litvak V, Cagnan H, Price CJ, et al. (2019): A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *Neuroimage*. 200:12-25.

28. Rosa MJ, Friston KJ, Penny W (2012): Post-hoc selection of dynamic causal models. *J Neurosci Meth*. 208:66-78.

29. Friston KJ, Penny W (2011): Post hoc Bayesian model selection. *Neuroimage*. 56:2089-2099.

30. Spielberger CD State-Trait Anxiety Inventory. *The Corsini Encyclopedia of Psychology*, pp 1-1.

31. Esmenio S, Soares JM, Oliveira-Silva P, Goncalves OF, Friston K, Coutinho JF (2020): Changes in the Effective Connectivity of the Social Brain When Making Inferences About Close Others vs. the Self. *Frontiers in Human Neuroscience*. 14.

32. Davey CG, Harrison BJ (2022): The self on its axis: a framework for understanding depression. *Transl Psychiatry*. 12:23.

33. Burgess PW, Dumontheil I, Gilbert SJ (2007): The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn Sci*. 11:290-298.

34. Apps MAJ, Ramnani N (2017): Contributions of the Medial Prefrontal Cortex to Social Influence in Economic Decision-Making. *Cereb Cortex*. 27:4635-4648.

35. D'Argembeau A (2013): On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Front Hum Neurosci*. 7:372.

36. Clark DM, Wells A (1995): A cognitive model of social phobia. *Social phobia: Diagnosis, assessment, and treatment*. New York, NY, US: The Guilford Press, pp 69-93.

37. Boehme S, Miltner WH, Straube T (2015): Neural correlates of self-focused attention in social anxiety. *Soc Cogn Affect Neurosci*. 10:856-862.

38. Chambless DL, Tran GQ, Glass CR (1997): Predictors of response to cognitive-behavioral group therapy for social phobia. *J Anxiety Disord*. 11:221-240.

39. Cui Q, Vanman EJ, Long Z, Pang Y, Chen Y, Wang Y, et al. (2017): Social anxiety disorder exhibit impaired networks involved in self and theory of mind processing. *Soc Cogn Affect Neurosci*. 12:1284-1295.

40. Yuan C, Zhu H, Ren Z, Yuan M, Gao M, Zhang Y, et al. (2018): Precuneus-related regional and network functional deficits in social anxiety disorder: A resting-state functional MRI study. *Compr Psychiatry*. 82:22-29.

41. Imperatori C, Farina B, Adenzato M, Valenti EM, Murgia C, Marca GD, et al. (2019): Default mode network alterations in individuals with high-trait-anxiety: An EEG functional connectivity study. *J Affect Disord*. 246:611-618.

42. Angeletti LL, Scalabrini A, Ricca V, Northoff G (2021): Topography of the Anxious Self: Abnormal Rest-Task Modulation in Social Anxiety Disorder. *Neuroscientist*.10738584211030497.

43. Mah L, Szabuniewicz C, Fiocco AJ (2016): Can anxiety damage the brain? *Curr Opin Psychiatr*. 29:56-63.

44. Paus T (2005): Mapping brain maturation and cognitive development during adolescence. *Trends Cogn Sci*. 9:60-68.

45. Blakemore S-J, den Ouden H, Choudhury S, Frith C (2007): Adolescent development of the neural circuitry for thinking about intentions.

46. Sebastian C, Burnett S, Blakemore SJ (2008): Development of the self-concept during adolescence. *Trends Cogn Sci*. 12:441-446.

47. Dixon ML, Moodie CA, Goldin PR, Farb N, Heimberg RG, Zhang JX, et al. (2022): Frontoparietal and Default Mode Network Contributions to Self-Referential Processing in Social Anxiety Disorder. *Cogn Affect Behav Ne*. 22:187-198.

Tables

Table 1.
Demographic and Behavioral Differences Between Healthy Controls and Social Anxiety Disorder Participants

	Healthy Controls (*N* = 72)		SAD participants (*N* = 38)			
Characteristics	*N* or Mean	Percentage or *SD*	*N* or Mean	Percentage or *SD*	Cramer's *V* or Hedges' *g*	*P*
Female	45	62.5	23	60.5	.03	.826
Comorbid MDD	-	-	33	86.8	-	-
Age (years)	20.5	1.8	19.8	2.3	.35	.166
Standardized IQ	111.9	8.4	110.3	13.4	.15	.816
Liebowitz Social Anxiety Scale score	23.2	17.1	84.5	24.7	3.06	<.001
State-Trait Anxiety Inventory General Score	33.1	8.9	57.4	9.1	9.04	<.001
State-Trait Anxiety Inventory Present Score	28.9	7.1	43.5	8.5	7.71	<.001
RT for external attention (seconds)	2.1	.47	2.1	.43	.03	.826
Accuracy for external attention (%)	94.4	5.8	94.0	5.7	.07	.868
Reaction time for direct self-appraisal (seconds)	1.8	.31	1.8	.31	.23	.816
Reaction time for reflected self-appraisal (seconds)	1.7	.30	1.9	.32	.64	.068
Proportion of adjectives affirmed during direct self-appraisal (%)	45.6	10.8	51.8	12.7	.53	.008
Proportion of adjectives affirmed during reflected self-appraisal (%)	42.7	12.2	47.9	14.3	.40	.050

Figures

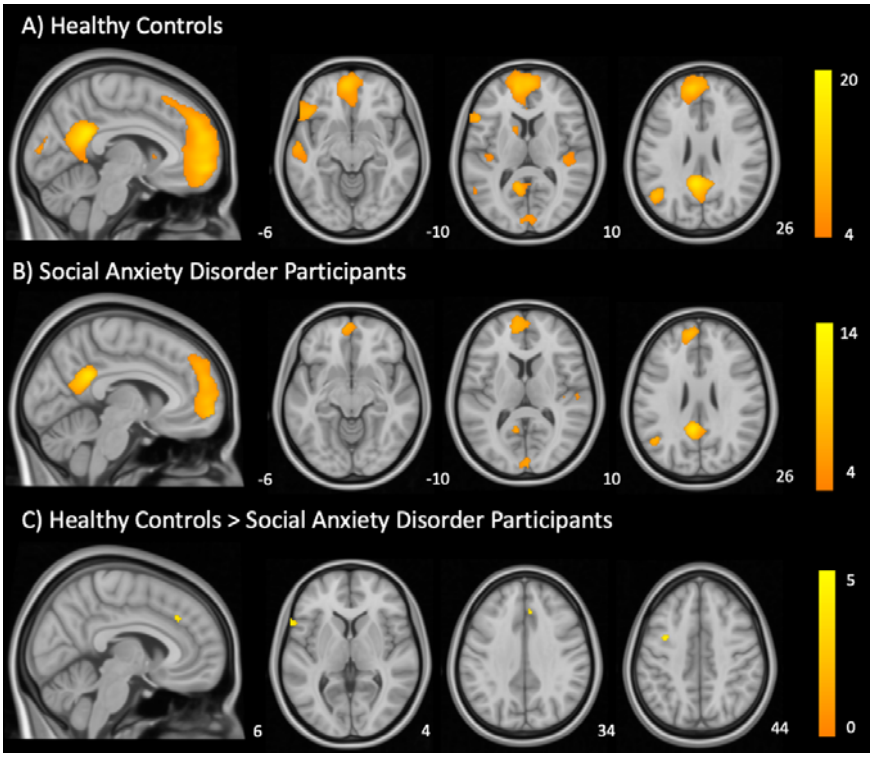

Figure 1. Regions of significant activation during direct self-appraisal compared with external attention for (A) healthy control participants and (B) SAD participants. Results are thresholded at $P_{FWE} < .05$, $K_e = 20$. Also shown is (C) the comparison of healthy controls and SAD participants, which is thresholded at $P < 0.001$, uncorrected, $K_e = 20$.

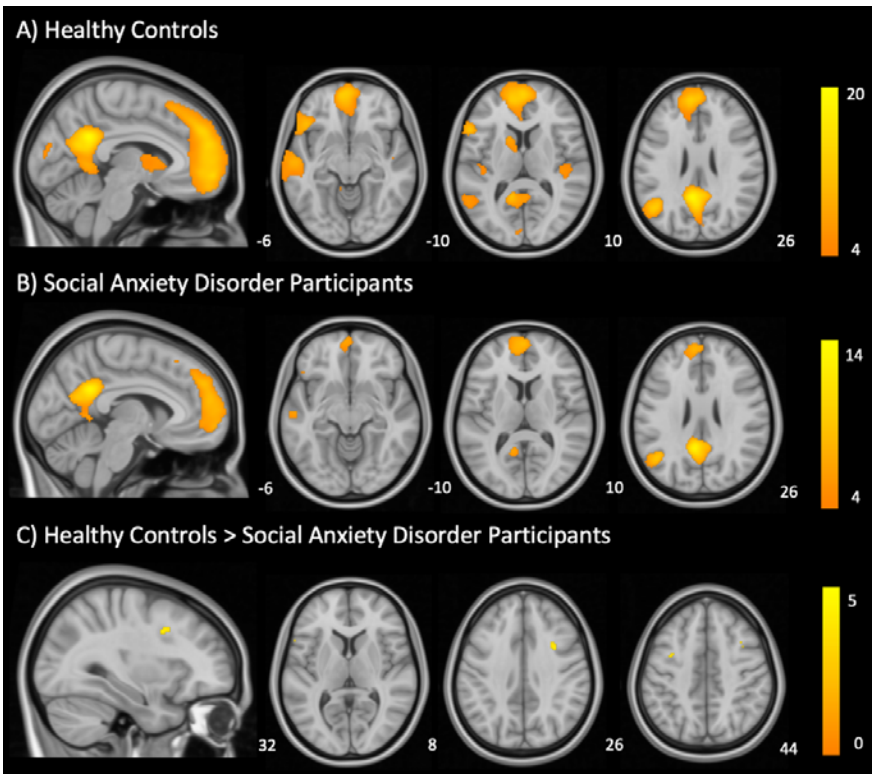

Figure 2. Regions of significant activation during reflected self-appraisal compared with external attention for (A) healthy control participants and (B) SAD participants. Results are thresholded at $P_{FWE} < .05$, $K_e = 20$. Also shown is (C) the comparison of healthy controls and SAD participants, which is thresholded at $P < 0.001$, uncorrected, $K_e = 20$.

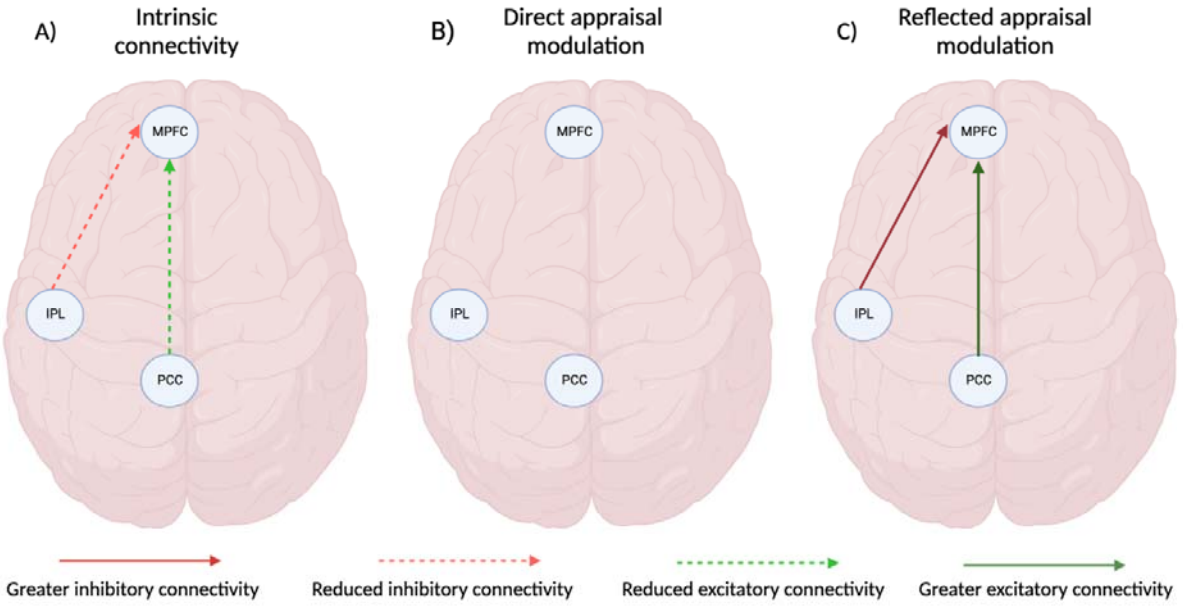

Figure 3. Differences between SAD participants and health controls in the intrinsic connectivity and modulation due to modulation associated with direct and reflected self-appraisal (PP > .95). Image created with BioRender (www.biorender.com).

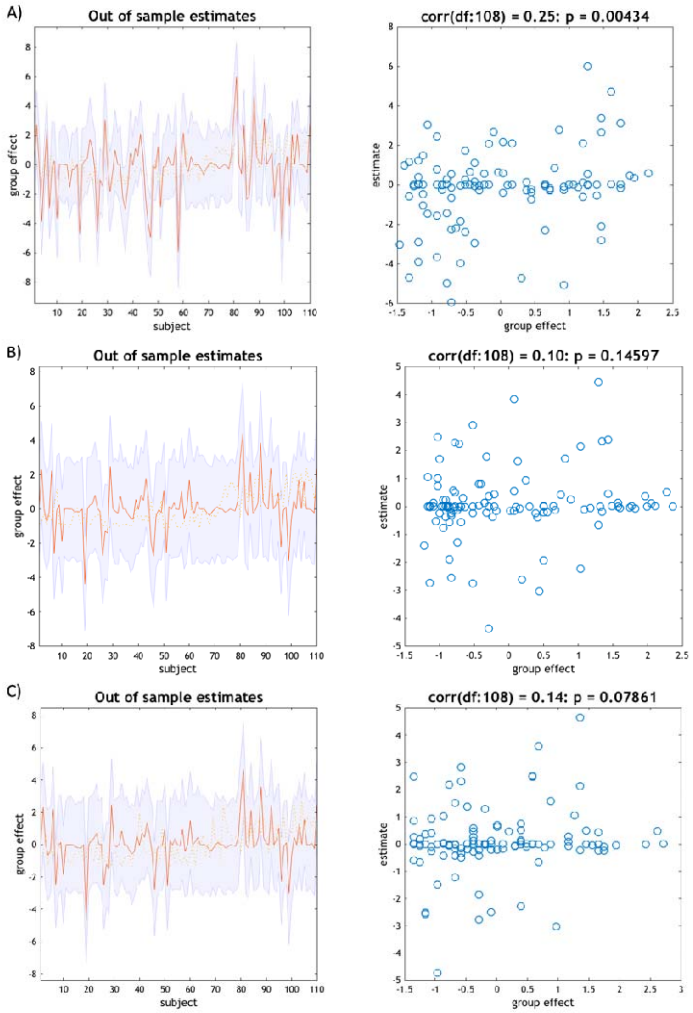

Figure 4. Leave-one-out cross-validation predicting anxiety symptom severity across all participants. **Left:** The out-of-samples estimated of anxiety symptom severity (standardized mean-centered) with 90% credible interval (shaded area). **Right:** The correlation between observed scores and the expected values for the (A) State-Trait Anxiety Inventory General Subscale, (B) Liebowitz Social Anxiety Scale, and (C) State-Trait Anxiety Inventory Present Subscale.