

Original Article

Cite this article: Shany O, Dunsky N, Gilam G, Greental A, Gilboa-Schechtman E, Hendler T (2022). Self-evaluation of social-rank in socially anxious individuals associates with enhanced striatal reward function. *Psychological Medicine* 1–11. <https://doi.org/10.1017/S0033291722001453>

Received: 19 August 2021

Revised: 21 April 2022

Accepted: 4 May 2022

Key words:

Emotion; fMRI; Reward; Self-evaluation; Social anxiety

Authors for correspondence:


Ofir Shany,

E-mail: ofirshany@gmail.com

Talma Hendler,

E-mail: talma@tlvmc.gov.il

Self-evaluation of social-rank in socially anxious individuals associates with enhanced striatal reward function

Ofir Shany^{1,2} , Netta Dunsky^{2,3}, Gadi Gilam⁴, Ayam Greental^{2,3},
Eva Gilboa-Schechtman⁵ and Talma Hendler^{1,2,3,6}

¹School of Psychological Sciences, Tel-Aviv University, Tel-Aviv, Israel; ²Sagol Brain Institute, Tel-Aviv Sourasky Medical Center, Tel-Aviv, Israel; ³Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel; ⁴Faculty of Dental Medicine, The Institute of Biomedical and Oral Research, Hebrew University of Jerusalem, Jerusalem, Israel; ⁵Department of Psychology and the Gonda Brain Science Center, Bar-Ilan University, Ramat-Gan, Israel and ⁶Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

Abstract

Background. Negative self-views, especially in the domain of power (i.e. social-rank), characterize social anxiety (SA). Neuroimaging studies on self-evaluations in SA have mainly focused on subcortical threat processing systems. Yet, self-evaluation may concurrently invoke diverse affective processing, as motivational systems related to desired self-views may also be activated. To investigate the conflictual nature that may accompany self-evaluation of certain social domains in SA, we examined brain activity related to both threat and reward processing.

Methods. Participants ($N = 74$) differing in self-reported SA-severity underwent fMRI while completing a self-evaluation task, wherein they judged the self-descriptiveness of high- *v.* low-intensity traits in the domains of power and affiliation (i.e. social connectedness). Participants also completed two auxiliary fMRI tasks designated to evoke reward- and threat-related activations in the ventral striatum (VS) and amygdala, respectively. We hypothesized that self-evaluations in SA, particularly in the domain of power, involve aberrant brain activity related to both threat and reward processing.

Results. SA-severity was more negatively associated with power than with affiliation self-evaluations. During self-evaluative judgment of high-power (e.g. dominant), SA-severity associated with increased activity in the VS and ventromedial prefrontal cortex. Moreover, SA-severity correlated with higher similarity between brain activity patterns activated by high-power traits and patterns activated by incentive salience (i.e. reward anticipation) in the VS during the reward task.

Conclusions. Our findings indicate that self-evaluation of high-power in SA involves excessive striatal reward-related activation, and pinpoint the downregulation of VS-VMPFC activity within such self-evaluative context as a potential neural outcome for therapeutic interventions.

Introduction

Social anxiety (SA) is a widespread affective disturbance that is marked by debilitating fear and avoidance of social situations (Leichsenring & Leweke, 2017). SA is associated with erosion of positivity in self-evaluations with regards to specific domains (Gilboa-Schechtman, Keshet, Peschard, & Azoulay, 2020). Therefore, it is crucial to elucidate the affective processes that accompany non-positive self-evaluations in SA. Based on social-cognitive perspectives on self-processing and clinical theories of SA, here we test the possibility that specific content of self-views may expose relations between SA and affective brain activity during non-positive self-evaluation.

Negative self-evaluation in SA may center around specific social domains (Gilboa-Schechtman et al., 2020; Moscovitch, 2009). One conceptualization of such domains derives from the evolutionary perspective on SA (Gilbert, 2001). This viewpoint suggests that high-SA individuals are sensitive to cues denoting social *power* (aka social-rank, status, dominance), and perceive themselves as possessing low power. Moreover, excessive attunement to power-related cues often comes at the expense of attentiveness towards signals of *affiliation* (i.e. social connectedness). Respectively, high SA individuals assign enhanced significance to power-related traits and evaluate themselves more negatively on power-related *v.* affiliative traits (Aderka, Weisman, Shahar, & Gilboa-Schechtman, 2009; Berger, Keshet, & Gilboa-Schechtman, 2017; Dijk, van Emmerik, & Grasman, 2018; Gilboa-Schechtman et al., 2017; Kashdan, Elhai, & Breen, 2008; Roberts, Hart, Coroiu, & Heimberg, 2011; Weisman, Aderka, Marom, Hermesh, & Gilboa-Schechtman, 2011).

Two prominent social-cognitive theories have described possible connections between self-evaluations and affective processing. The theory of ‘possible selves’ (Markus & Nurius, 1986) states that mental representations of what individuals could hopefully become, or of what they

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE
UNIVERSITY PRESS

are afraid of becoming, are cognitive manifestations of underlying desires or threats, respectively. The 'self-discrepancy' theory (Higgins, 1987) posits that incompatibility between desired and actual self-views may elicit negative emotions and dampen positive emotions. These theories both suggest that pondering desired (but not actual) and non-desired traits, and evaluating these traits against the current self, mobilizes affective processes that are likely mediated through dedicated neural systems. Specifically, self-evaluation may encompass signaling valuable self-concepts, a process that is mediated in part through the brain's reward circuit (Bartra, McGuire, & Kable, 2013; D'Argembeau et al., 2012). Furthermore, the affirmation of positive or negative qualities of the self, may arise pleasant and unpleasant feelings that scale with reward- and threat-related brain activity (Eisenberger, Inagaki, Muscatell, Haltom, & Leary, 2011; Izuma, Kennedy, Fitzjohn, Sedikides, & Shibata, 2018). High-SA individuals view power-related traits as important to their self-concept, yet they are largely discrepant from their actual self-view (Gilboa-Schechtman, Galili, Sahar, & Amir, 2014; Kashdan et al., 2008; Roberts et al., 2011). Thus, evaluation of the self in terms of power may concurrently involve altered activations in both reward- and threat-related neural circuitries in SA. Neuroimaging provides a powerful means for associating affective neural processing with self-evaluation.

Hitherto, neuroimaging studies have associated self-referential processing with cortical midline components of the default-mode network (DMN). These include dorsal and ventral portions of the medial prefrontal cortex (MPFC), as well as the precuneus and posterior cingulate cortex (PCC). The amalgamation of self-referential processing with affect likely involves subcortical regions. These include the amygdala, which is implicated in negative affective processing, including threat to the integrity of one's self-view (Frewen, Lundberg, Brimson-Théberge, & Théberge, 2013; Hughes & Beer, 2013); and the VS – a major node of the mesolimbic reward system. The VS mediates processes such as incentive salience (i.e. signaling motivationally significant stimuli); and reward responsiveness (i.e. consuming rewarding outcomes) (Bartra et al., 2013; Liu, Hairston, Schrier, & Fan, 2011). In context of self-evaluation, VS activation has been linked to processing important self-related content such as one's reputation (Izuma, Saito, & Sadato, 2010), and to affirming positive qualities of oneself (Chavez & Heatherton, 2014; Izuma et al., 2018). Given the involvement of amygdala and VS in self-evaluation, an examination of self-evaluation in SA should consider both of these affect-related neural regions.

In clinically diagnosed SA, self-referential processing increases activation in midline DMN structures and in the amygdala relative to healthy controls (Blair et al., 2008, 2011; Dixon et al., 2020; Goldin & Gross, 2010; Goldin, Manber-Ball, Werner, Heimberg, & Gross, 2009b). Moreover, while increased MPFC activity is associated with tracking the self-descriptiveness of traits (D'Argembeau, 2013; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006), in SA it has been implicated in processing both negative and positive (and thus probably less self-descriptive) self-related information such as traits and social feedbacks (Goldin, Ramel, & Gross, 2009a; Heitmann et al., 2014; Peterburs, Sandrock, Miltner, & Straube, 2016). These findings concord with the enhanced self-focused attention and negative affect that characterize SA (Norton & Abbott, 2016). Yet, these studies mostly utilized personalized negative self-beliefs or valenced traits without specifying their social domain. Moreover, while these studies often focused on threat-related amygdala activity, engaging reward-related activation in the VS may also be crucial

for positive self-evaluation as discussed above. Several studies show aberrant VS functionality in SA, including increased response to interpersonal situations and threatening social cues (Finlayson-Short, Harrison, & Davey, 2021; Klumpp, Angstadt, & Phan, 2012), and reduced response to social rewards (Becker, Simon, Miltner, & Straube, 2017; Schultz et al., 2019). However, the role of reward-related VS activity in poor self-evaluation in SA remains unclear. Moreover, the amygdala and VS are not exclusively dedicated to the processing of negative and positive valenced information, respectively. Rather, both regions are associated with diverse affective and motivational processes (Bartra et al., 2013; Putnam & Gothard, 2019). Thus, *which* affective processes are actually mediated by these subcortical regions during biased self-evaluation in SA demands clarification.

The current study aimed to concurrently explore reward- and threat-related brain activations that accompany self-evaluative processing in SA, while taking into account specific content dimensions of self descriptions – namely power and affiliation. Providing a more nuanced understanding of the content of concerns in SA and their underlying neural mechanisms, could inform both cognitive-behavioral- and neuromodulation-based therapeutic approaches to SA. To this end, participants ranging in SA-severity completed a self-referential encoding task (SRET) during an fMRI scan. Participants were requested to decide on the self-descriptiveness of traits derived from domains of power and affiliation. Traits differed in their denomination of high and low end of the power and affiliation continua with some traits denoting high end of the continuum (e.g. assertive, warm) and some low end of the continua (weak, hostile). To characterize the affective processes that are mediated by subcortical regions during the SRET, participants completed two additional auxiliary fMRI tasks that are known to evoke VS and amygdala activity in response to reward and threat, respectively. These included a monetary reward task, which probed participants' anticipation (i.e. incentive salience phase) and consumption (i.e. reward responsiveness phase) of monetary prizes; and a threat task, wherein participants executed a simple matching task on angry and fearful faces.

Two hypotheses were tested. First, we expected SA-severity to be more negatively associated with power than with affiliation self-evaluations (Gilboa-Schechtman et al., 2020). We hypothesized that this pattern would manifest at the neural level, such that SA-severity will be associated with greater activity in midline DMN nodes (MPFC, precuneus-PCC) and in the amygdala (Brühl, Delsignore, Komossa, & Weidt, 2014; Yoon, Seo, Kim, & Choo, 2019), during the self-evaluation of power (*v.* affiliation) traits. Second, we hypothesized that reflecting on power (*v.* affiliation) in higher SA-severity would associate with reward-related VS activity, and our hypothesis regarding this matter was two-sided. On the one hand, insufficiently positive self-evaluation in terms of power in higher SA-severity may be accompanied by *diminished* VS activity. On the other hand, SA is associated with assignment of high value to power-related traits (Kashdan et al., 2008; Roberts et al., 2011), which may be reflected by *increased* VS activity (Bartra et al., 2013). Since reward-related effects might manifest particularly with regards to more positive traits (Chavez & Heatherton, 2014) such as the high-intensity traits in this study (see online Supplementary Methods), we also examined whether brain activation in the SRET was affected by the interaction of SA-severity with both the social domain and intensity of traits. Lastly, in order to better characterize affective processes that were possibly engaged during self-evaluation of power in SA, we implemented a neural representational similarity

analysis between the SRET and the auxiliary tasks. This analytical approach can reveal similarities in how the brain encodes affective and cognitive processes across tasks and mental states. We reasoned that greater neural similarity between the SRET and the reward and threat tasks in a given brain region, would provide evidence that a region mediated the latter affective processes at the time of self-evaluation (Axelrod, Rees, & Bar, 2017; Chavez, Heatherton, & Wagner, 2017; Puccetti et al., 2021). Thus, in an exploratory analysis that followed the brain activation analysis, we examined the correlation of SA with between-task similarity indices in regions whose activation was modulated as a function of SA during self-evaluation of power in the SRET.

Methods

Participants

We recruited seventy-four participants ($M_{\text{age}} \pm \text{s.d.}$: 25.11 \pm 3.03 years, 44 females) to this experiment following an online screening procedure ($n = 845$; see online Supplementary Methods and Fig. S1), whose purpose was to assure the recruitment of participants ranging in SA-severity as indicated by the Liebowitz Social Anxiety Scale (LSAS) (Liebowitz, 1987). The sampling strategy was a convenience/availability sampling, meaning that participants signed up to participate and were included if they fulfilled the inclusion criteria (detailed in online Supplementary Methods). Moreover, we did not conduct any formal power analysis to determine the fMRI sample size, but it is consistent with literature recommendations (Durnez et al., 2016; Turner, Paul, Miller, & Barbey, 2018) and with similar recent fMRI studies that focused on brain-behavior correlations (Will et al., 2020; Yoon, Somerville, & Kim, 2018). All participants completed the SRET; 71 completed the threat task ($M_{\text{age}} \pm \text{s.d.}$: 25.10 \pm 3.07 years, 42 females); and 54 participants completed the reward task ($M_{\text{age}} \pm \text{s.d.}$: 25.15 \pm 3.11 years, 33 females). Note that the first 20 participants completed a different reward task that involved risk taking in a social context, and we decided to replace it with a simpler and shorter task. Three/two/four participants were excluded from the fMRI analysis of the SRET/reward/threat tasks, respectively, due to technical issues and exaggerated head movement (see online Supplementary Methods). Thus, fMRI analysis was performed on 71 participants in the SRET ($M_{\text{age}} \pm \text{s.d.}$: 25 \pm 2.96 years, 43 females); 52 participants in the reward task ($M_{\text{age}} \pm \text{s.d.}$: 25.19 \pm 3.1 years, 31 females); and 67 participants in the threat task ($M_{\text{age}} \pm \text{s.d.}$: 25 \pm 2.98 years, 41 females).

Procedure overview

At 1–2 days prior to the experiment, participants filled the LSAS questionnaire again via the *qualtrics* software. On the experimental day, participants signed an informed consent and practiced the three fMRI tasks (15–20 min). Next, they entered the MRI scanner, and completed an anatomical scan and the SRET (2 runs, ~5 min each), reward (2 runs, ~5 min each) and threat tasks (~7 min). To control for potential effects of reward on subsequent threat processing or vice versa, presentation order of the reward and threat tasks was counter-balanced across participants.

Liebowitz Social Anxiety Scale (LSAS)

The LSAS is a 24-item scale measuring fear and avoidance of social situations over the past week. Each item is rated on two

4-point Likert scales that measure fear and avoidance regarding the item, from 0 (never) to 3 (severe). SA ranges in severity, but even below-diagnostic levels are associated with reduced well-being (Fehm, Beesdo, Jacobi, & Fiedler, 2008) and impairments in work performance and social relationships, which are akin to those evident in full-blown social anxiety disorder (SAD) (Wittchen, Fuetsch, Sonntag, Müller, & Liebowitz, 2000). Respectively, dimensional approaches to SAD suggest that its diagnosis and clinical outcomes are best explained by viewing this condition as existing along a continuum that ranges from subclinical levels and up to full syndrome levels, rather than as a dichotomous condition that is either present or absent (Ruscio, 2010). Crucially, dimensional approaches advocate for examination of subclinical levels of SA, as this may enhance accurate prognosis by identifying risk factors of this disorder (e.g. genetic), as well as guide its effective treatment (Ruscio, 2010). Therefore, we aimed to sample SA in a dimensional manner (Hyett & McEvoy, 2018) that would span from low levels of SA (LSAS \leq 30) and up to clinical levels that characterize SAD (LSAS \geq 50) and generalized SAD [LSAS \geq 60; (Mennin et al., 2002; Rytwinski et al., 2009). Indeed, our sample included a wide range of SA-severity (LSAS scores at screening ranged between 8–109; $M \pm \text{s.d.}$: 51.35 \pm 23.07). Adjacent to the experimental day (i.e. during which participants filled the LSAS questionnaire for the second time), LSAS scores ranged between 2–108 ($M \pm \text{s.d.}$: 51.15 \pm 26.12; see LSAS distributions in online Supplementary Fig. S1 and Supplementary Methods for more details). The correlation between the two LSAS measurements was high ($r(72) = 0.86$, $p < 0.001$), and for the analyzes we used LSAS scores obtained on the experimental day.

fMRI tasks

Self-Referential Encoding Task (SRET)

Neurobehavioral processing of self-views was assessed with a modified version of the SRET (Goldin, Ziv, Jazaieri, & Gross, 2012). Throughout the task, participants were exposed to the same set of traits twice. On each presentation they either judged if the traits were self-descriptive on ‘self’ blocks; or determined whether the first 2 letters of the trait were presented in the alphabetical order on ‘control’ blocks (Axelrod et al., 2017). The stimuli consisted of 64 social trait adjectives (online Supplementary Table S1), which were classified into a 2 \times 2 within-subject factorial design according to their domain (power *v.* affiliation) and intensity (high *v.* low loading of the traits on the relevant dimensions). Domain and intensity of stimuli were validated in an independent behavioral experiment (see online Supplementary Methods and Tables S1 and S2).

The SRET was conducted using an fMRI block design (Fig. 1). Each block initiated with a fixation (2.5s) that was followed by a task instruction screen (2.5s). Afterwards, a set of 4 traits pertaining to one of the experimental conditions was presented (12.5s; 3.125s per trait). Participants used an MRI-compatible response box to provide yes/no answers while each trait appeared. The task was performed in 2 separate runs (~5 min each), and each run contained 16 blocks and started with a 22.5s long fixation. Stimulus order included a pseudo-random block sequence with no more than two consecutive blocks of the same condition and a random sequence of words within each block. Presentation order of conditions was counterbalanced between participants, by using four

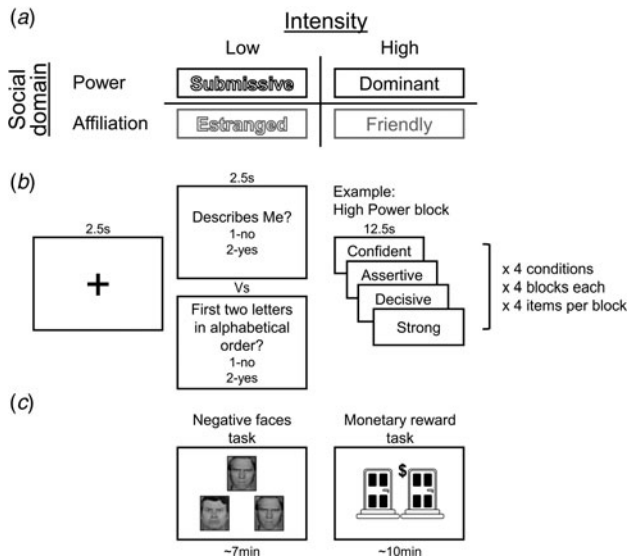


Fig. 1. Self-referential encoding task (SRET) and experimental design. (a). The traits in the SRET were classified into four experimental conditions according to their social domain (power v. affiliation; colored in black v. gray, respectively) and intensity (low v. high; without or with color fill, respectively), thus yielding a 2×2 within-subject factorial design. (b). Example of an experimental block in the high-power condition. (c). Following the SRET, the participants completed two auxiliary fMRI tasks assessing brain activity in response to reward- and threat-related processes. In the threat task, participants performed a simple matching task on negative facial expressions (angry and fearful) or geometrical shapes (control condition). On each block, participants were asked to select one of two faces/shapes (located at the bottom right or bottom left of the screen) that matched the target face/shape (located at the top of the screen). In the monetary reward task, on each trial participants were asked to select one of two doors, which, following an anticipatory phase, led to either monetary prizes or losses. More details on these tasks are presented in the online Supplementary Methods and in Figs S2 and S3. Note that administration order of the two auxiliary fMRI tasks was counter-balanced across participants.

different task versions. The experiment was programmed and presented in PsychoPy (version 1.85).

Auxiliary fMRI tasks for characterizing basic affective processes

The auxiliary tasks' structures are fully detailed in the online Supplementary Methods and Figs S2 and S3. Reward-related processes in the VS were probed via a gambling task (Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011). On each trial of this event-related paradigm, participants had to select one of two doors (2.5s), which, following an anticipatory phase (2.5s), led to either monetary prizes (+5 NIS) or losses (−2.5 NIS) that were signified by an upward-pointing green arrow or a downward-pointing red arrow, respectively (1s). Unknown to participants, the task included an equal amount of predetermined wins and losses (30 trials each) that appeared in a fixed order. Stimulation of threat-related activation in the amygdala was probed through the emotional faces matching task (Hariri, Tessitore, Mattay, Fera, & Weinberger, 2002; Scult, Knodt, Radtke, Brigidi, & Hariri, 2019). In this block-design task, participants performed either a face-matching task on emotional facial expressions (angry, fearful, surprised, and neutral, presented on 4 different blocks) or a similar sensorimotor control task on geometrical shapes (5 blocks). Specifically, participants were instructed to select one of two faces/shapes (located at the bottom right or bottom left of the screen) that matched the target face/

shape (located at the top of the screen). Faces block lasted 48s and shapes blocks lasted 36s. Each block consisted of 6 presentations of faces/shapes trios the lasted 4s, interleaved with fixations. Here we analyzed only angry and fearful faces, as they convey negative affect directly.

Behavioral data analysis

Interaction of SA with endorsement of power v. affiliation traits

To examine the endorsement of traits in the SRET, we first computed the percentage of traits that participants endorsed as self-descriptive per condition. Note that due to a technical error that occurred in 40 participants in the high-power condition, one block was removed from each condition in all analyses (see online Supplementary Methods). To examine our hypothesis that SA-severity is negatively associated with endorsement of power traits above and beyond the association with affiliation traits, we tested the interaction between LSAS scores and within-subject categorical factors (domain and intensity) in predicting endorsement of traits. We conducted a linear moderated regression analysis using the *pequod* package in R (<https://CRAN.R-project.org/package=pequod>), whose model included predictors for all main and interaction effects. Endorsement percentages of low-intensity traits were reverse-scored by subtracting them from 100, such that they would represent higher self-descriptiveness within their social domain. Interaction effects involving LSAS scores were inspected by computing simple slopes and testing their differences using the *pequod* package as well (Dawson & Richter, 2006).

Functional Magnetic Resonance Imaging (fMRI)

Imaging data acquisition

See online Supplementary Methods.

Imaging data preprocessing

Imaging data were preprocessed with *fMRIPrep* (Esteban et al., 2019). Details are provided in the online Supplementary Methods. Note that data of two/four participants were excluded from the analysis of the reward/threat tasks, respectively, due to head movements exceeding 3 mm.

Analysis of fMRI data from the SRET

Statistical analysis of the fMRI data was conducted with Statistical Parametric Mapping software (SPM12; <http://www.fil.ion.ucl.ac.uk/spm>). The 1st-level General Linear Model (GLM) of the task included 10 predictors: (1–8) predictors for the 8 task conditions (event duration = 12.5s); (9–10) predictors for self and control blocks that were excluded from the analysis (see online Supplementary Methods). All predictors were convolved with a canonical hemodynamic response function, and data were subjected to SPM12 default high-pass filter cutoff (128s). Confound regressors were added to the 1st-level model as well (see online Supplementary Methods). We then computed linear contrasts for the effects of social domain and for the intensity by social domain interaction for each participant. Note that within these contrasts, we also subtracted brain activity measured during the control condition blocks from their corresponding self condition blocks (e.g. from the high-power self condition we subtracted the high-power control condition). Next, we entered these contrasts to two 2nd-level multiple regressions with LSAS scores as a covariate. The resulting activation maps were thresholded with a voxel-

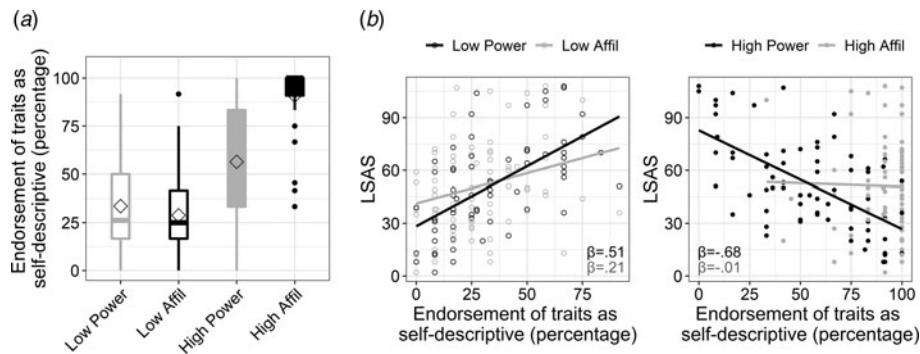


Fig. 2. Endorsement of traits according to their social domain and intensity, and its interaction with social anxiety. (a). Boxplots depicting the percentage of endorsing traits as self-descriptive (i.e. responding ‘yes’ to the question ‘does the trait describe me?’) in each of the four experimental conditions. The power *v.* affiliation (abbreviated as affil) social domains are colored in black and gray, respectively. High *v.* low intensity is denoted by color fill or empty fill, respectively. The horizontal line in each boxplot marks the median, and upper and lower bounds of the box denote the 75th (Q3) and 25th (Q1) percentiles, respectively. The line stretches between the minimum (Q1 – 1.5 × interquartile range) and maximum (Q3 + 1.5 × interquartile range) of the data. The diamond shape indicates the mean. (b). Exploration of the 3-way interaction of social anxiety level, as measured by the Liebowitz Social Anxiety Scale (LSAS), with intensity and social domain of traits. The interaction was explored by computing and comparing between simple slopes that quantified the association of LSAS (*y*-axis) with endorsing traits within each condition. The simple slopes for LSAS (β -value) within each of the traits’ conditions are overlaid on each scatterplot and are colored according to the social domain.

level threshold of $p < 0.001$ (Eklund, Nichols, & Knutsson, 2016; Woo, Krishnan, & Wager, 2014) and cluster-level correction at $p_{FWE} < 0.05$. For basic validation of the SRET, we computed a contrast estimating the effect of self *v.* control (i.e. summation of self *v.* control blocks) and subjected these contrasts to a group-level one-sample *t* test. Group-level effects of the tasks conditions were inspected as well in exploratory analyses.

Region of Interest (ROI) Representational Similarity Analysis (RSA)

In order to better interpret the underlying affective processes of self-evaluation that were mediated by subcortical regions as a function of SA, we conducted an exploratory RSA between the SRET and each of the auxiliary tasks. We assumed that higher between-task similarity would indicate to some degree that participants engaged reward- or threat-related neural representations in a given ROI when self-evaluating (Chavez et al., 2017). In the SRET, we focused our analysis on the high-power self > high-power control contrast, due to association of LSAS with both negative self-evaluation and altered brain activity in response to this condition (see Results). We extracted the following contrasts from the auxiliary tasks (see online Supplementary Methods for details on GLM construction in these tasks): from the reward task, we extracted one contrast reflecting reward responsivity (winning > losing money), and another contrast capturing incentive salience (anticipation > implicit baseline). From the threat task, we extracted one contrast representing sensitivity to social threat cues (fearful + angry faces > shapes; hereinafter termed ‘negative faces’). We focused our analysis on a right VS and left VMPFC ROIs [regions no. 220 and 47 in the Brainnetome atlas; (Fan et al., 2016)], since LSAS correlated with response to high-power in these regions (see Results). We then correlated the similarity values with LSAS scores, and controlled for multiple comparisons by applying a $p_{FDR} \leq 0.05$ threshold on the emerging results for each ROI.

Since RSA is a correlation-based measure and thus sensitive to the number of voxels, we wanted to assure that a similar amount of voxels was available for all participants. Thus, we re-created the ROIs based on their overlap with the

grouped-average brain mask produced during the 2nd-level analysis. The right VS contained 162 voxels and the left VMPFC contained 130 voxels. This analysis was carried with the RSA toolbox (Nili et al., 2014), and its computation pipeline is schematically depicted in Fig. 4a.

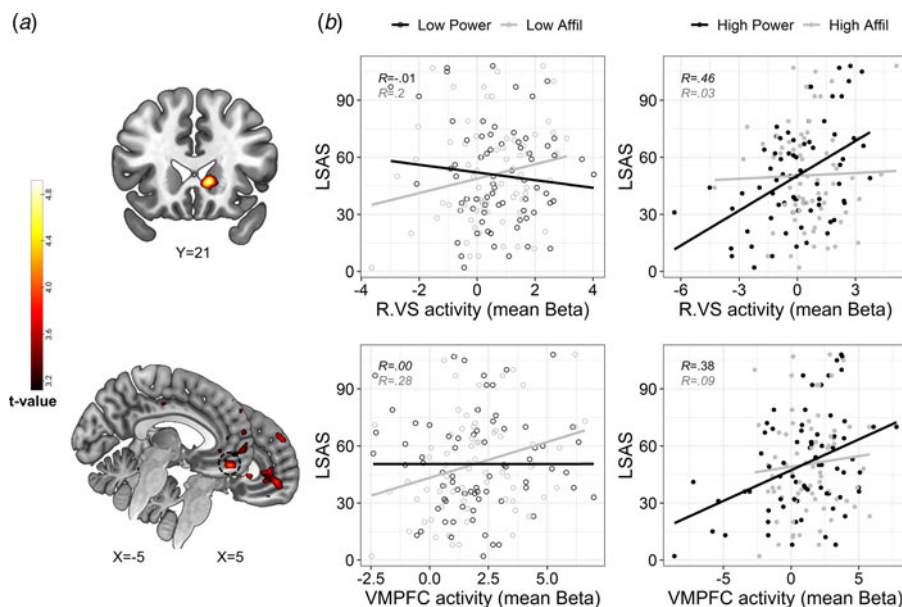
Results

Reduced endorsement of power *v.* affiliation traits in SA

We first examined whether the domain and intensity of traits in the SRET, as well as their interaction with LSAS scores, predicted the percentages of endorsing traits as self-descriptive. This linear regression was significant ($R^2 = 0.45$, Adjusted $R^2 = 0.43$, $F_{(7,288)} = 33.39$, $p < 0.001$). At the group-level, there were significant effects for social domain and its interaction with intensity on endorsement percentages, when LSAS was not included in the model (i.e. via repeated-measures ANOVA). These effects were mostly driven by the high endorsement of high-affiliation traits (Fig. 2a; see online Supplementary Tables S4–S6 for regression and repeated-measures ANOVA results, and online Supplementary Table S3 for descriptive statistics of the SRET).

In accordance with our hypothesis, SA-severity associated negatively with self-evaluation in terms of power [simple slope: $t(288) = -7.98$, $p < 0.001$] rather than affiliation [simple slope: $t(288) = -0.13$, $p = 0.9$; two-way interaction between LSAS and social domain: $\beta = -0.83$, $t = -5.55$, $p < 0.001$]. The analysis also revealed a significant three-way interaction between LSAS, domain and intensity of traits ($\beta = 0.37$, $t = 2.13$, $p = 0.03$; Fig. 2b). Post-hoc slope tests indicated that SA-severity was associated with negative self-evaluation in terms of power relative to affiliation within each intensity loading, but this effect was particularly robust within high-intensity traits [slope differences test for high-power *v.* high-affiliation: $t(288) = -5.55$, $p < 0.001$; for low-power *v.* low-affiliation: $t(288) = -2.53$, $p = 0.011$; see simple slope coefficients in Fig. 2b]. These results highlight that the association of SA-severity with negative self-evaluation in the power *v.* affiliation domain was more pronounced within the high end of the power and affiliation continua. Note that effects of the SRET conditions on response time in the self

Fig. 3. Modulation of brain activity by the interaction between social anxiety, domain and trait intensity. (a). Statistical parametric maps resulting from a whole-brain analysis of the 3-way interaction. Significant interaction effects were revealed in clusters in the right VS (upper panel) and VMPFC (lower panel). Statistical threshold was set at a voxel-level $p < 0.001$ and cluster-level $p_{FWE} < 0.05$. Approximate location of a peak activation in the VMPFC that was used for post-hoc analysis is circled in a dashed black line in the lower panel. (b). Scatterplots depicting the correlation between SA-severity and brain activity during self-evaluation of low- and high-intensity traits (presented on the left v. right panels, respectively, with empty v. color fill) pertaining to affiliation v. power domains (colored in black v. gray, respectively), in the right VS and VMPFC (upper and lower panel, respectively). The plotted Beta values represent brain activity that was extracted from contrasting each self condition against its matching control condition. Linear trend lines and correlation coefficients (Pearson's R) are presented in each scatterplot and are colored according to their social domain, solely for illustrative purposes. Abbreviations: Right ventral striatum (R.VS); ventromedial prefrontal cortex (VMPFC); affiliation (affil). Brain images are presented in neurological convention (i.e. right is right).



conditions, as well on accuracy and response time in the control condition, are presented in the online Supplementary Results and Tables S7–S9.

fMRI results

Association of SA with high-power in VS and VMPFC

Turning to the fMRI data, we first verified that the SRET successfully replicated previous studies by associating self-referential v. control conditions with DMN activity. Notably, in the self-evaluation condition subcortical regions of interest (amygdala, VS) more strongly activated than in the control condition (online Supplementary Fig. S4). Group-level effects found for social domain and intensity of traits, as well as for the interaction of these factors, are presented in online Supplementary Table S10.

Next, we examined the hypothesis that SA-severity would be associated with altered DMN and subcortical activity specifically during self-referential processing of *power* traits. While the 2-way interaction between LSAS and social domain did not reveal significant effects, we found a significant 3-way interaction between LSAS scores, social domain, and intensity in the VMPFC and right VS – specifically the ventral caudate (Fig. 3a and online Supplementary Table S11). To explore the interaction, we extracted mean activity (beta value) from 4 mm spherical ROIs centered around peak activations in the VMPFC ($x = -5$, $y = 49$, $z = -19$) and right VS ($x = 11$, $y = 21$, $z = -1$) for each self-evaluation condition (v. control), and plotted these values against LSAS scores. In partial correspondence with our hypothesis, this post-hoc inspection revealed that SA-severity was associated with enhanced VMPFC activity specifically during the processing of high-power traits; whereas this pattern was not evident for high-affiliation traits. In addition, SA-severity was also positively associated with VS activity during self-evaluation of high-power (our hypothesis regarding this region was two-sided). Moderate or weak associations between SA-severity and brain activity were evident in the remaining conditions (Fig. 3b). Exploratory analyses

revealed that higher activity in the above-defined VMPFC and right VS spherical ROIs also associated with less endorsement of high-power traits [VMPFC: $r(69) = -0.26$, $p = 0.03$; right VS: $r(69) = -0.34$, $p = 0.004$]. However, the positive association with LSAS in the VMPFC and VS was only slightly attenuated after controlling for endorsement percentages [VMPFC: $r(68) = 0.29$; right VS: $r(68) = 0.34$; compare with Fig. 4b]. Thus, VMPFC and VS activity in response to high-power was uniquely correlated with SA-severity above and beyond the endorsement of these traits as self-descriptive.

Similarity between self-referential encoding of high-power and reward processing in VS associates with SA

We next tested if brain activity during self-referential processing, possibly mediated reward- or threat-related processes in correspondence with SA. We focused this analysis on high-power, due to the association of LSAS with both negative self-evaluation and altered brain activity in response to this condition; and on right VS and left VMPFC clusters wherein LSAS interacted with high-power (Fig. 3). We found that during self-referential encoding of high-power, individuals with higher SA-severity were more likely to activate a multivoxel pattern in the right VS that resembled a pattern evoked by incentive salience in the reward task [$r(50) = 0.33$, $p = 0.017$, $p_{FDR} = 0.05$, Fig. 4b]. Moreover, in the right VS ROI, LSAS was not related to similarity of high-power to either reward *responsiveness* in the reward task [$r(50) = -0.12$, $p = 0.39$, $p_{FDR} > 0.59$], nor to threat processing in the negative faces task [$r(65) = -0.01$, $p = 0.94$, $p_{FDR} > 0.94$]. Additional control analyses did not reveal similar correlations of LSAS with similarity between self-processing of traits and incentive salience in the remaining SRET conditions (online Supplementary Table S12); nor with endorsing high-power traits [$r(50) = -0.17$, $p = 0.22$]. For the left VMPFC ROI, no significant correlations were found between LSAS scores and similarity of the high-power contrast to any of the contrasts in the auxiliary tasks [incentive salience: $r(50) = 0.14$, $p = 0.33$; reward responsiveness: r

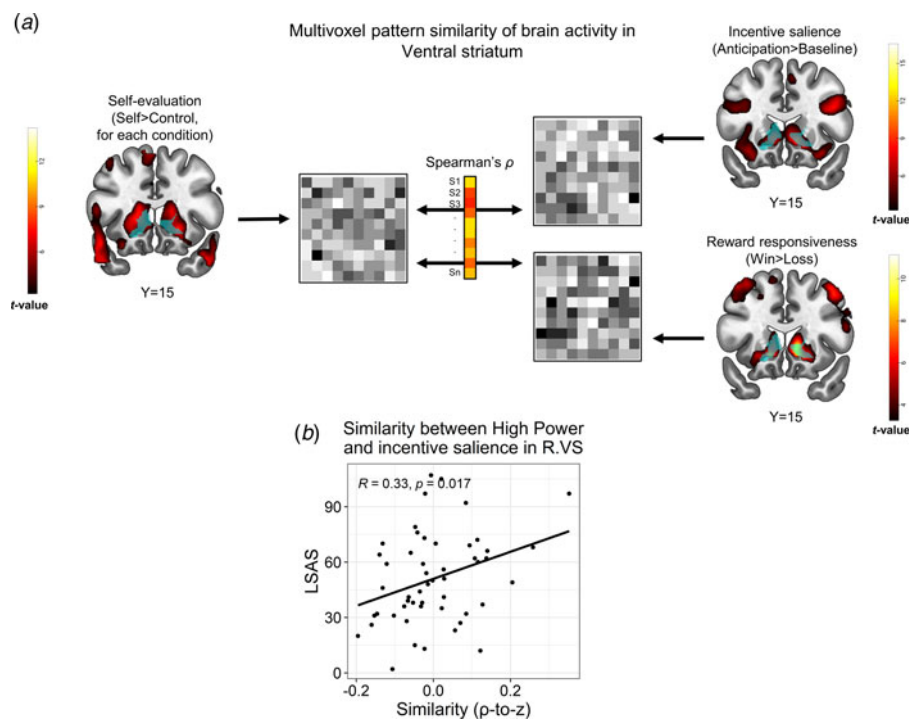


Fig. 4. Representational similarity analysis (RSA) between the self-referential encoding task and an auxiliary task probing reward processes. (a). Schema of the between-task RSA pipeline. Multivoxel brain activity patterns in a VS ROI (light blue) were extracted from a relevant contrast in the SRET (self > control for a specific condition; left panel), as well as from contrasts in the reward task that represented incentive salience processing (anticipation > baseline) and reward responsiveness (win > loss). Next, these patterns were correlated for each subject, thus yielding a Spearman's correlation coefficient, which represented the between-task similarity for that subject. We then applied a Fischer r -to- z transformation on the correlation coefficients and correlated them with LSAS scores across participants. (b). Correlations between LSAS scores and similarity of self-evaluation of high-power with incentive salience in the VS (right). Pearson's correlation coefficient and its corresponding p values are overlaid on the scatterplot. Abbreviations: right (R); ventral striatum (VS); Liebowitz social anxiety scale (LSAS). The statistical parametric maps show the group-level effect (i.e. t values of a one sample t test) obtained for each contrast (note that the map from the SRET shows the self > control contrast across all conditions) and are thresholded at voxel-level $p < 0.001$. All brain images are displayed in neurological convention.

(50) = -0.11 , $p = 0.44$; threat: $r(65) = -0.07$, $p = 0.58$]. Note that RSA analyses conducted at the whole-brain level revealed a significant group-level effect for similarity between self-evaluation (SRET) and incentive salience processing (reward task) within the VS (online Supplementary Fig. S5); and also a correlation between LSAS and similarity of high-power to incentive salience in the VS and VMPFC, albeit at an uncorrected statistical threshold (online Supplementary Fig. S6). These additional findings enhance the reliability of the between-task similarity analysis and corroborate the ROI RSA results.

Discussion

In this study, we examined the neural and affective processing of self-views in the domains of power and affiliation in SA. As expected, SA-severity was negatively associated with self-evaluation in the power domain, thereby corresponding to indications of impaired self-views in the domain of power *v.* relatively positive affiliation self-evaluation in clinical and subclinical SA (Berger et al., 2017; Gilboa-Schechtman et al., 2017). However, at the neural level, SA-severity associated with altered neural activity only during self-evaluation of high-power traits; specifically, with increased VS and VMPFC activity. Moreover, during self-processing of high-power traits, individuals with higher SA showed brain activity patterns that were more similar to those evoked by incentive salience in the VS.

Our finding that SA-severity was associated with increased activity in the VMPFC and right VS (ventral caudate) specifically during self-evaluation of high-power, both replicates and extends findings from previous examinations of negative self-evaluation in SA. Enhanced MPFC is recurrently evident during self-referential processing in clinically diagnosed SA (see Yoon et al., 2019 for a review). In specific, the foci of MPFC activations found here overlap with areas in the VMPFC and pregenual anterior cingulate cortex, which have been recurrently associated with processing

valuable self-relevant information (Bartra et al., 2013; D'Argembeau et al., 2012; Lieberman, Straccia, Meyer, Du, & Tan, 2019; Qin & Northoff, 2011). Similar areas were previously found hyper-activated in SA in response to both negative and positive self-related information such as self-beliefs and social feedbacks (Goldin et al., 2009a, 2009b; Heitmann et al., 2014; Peterburs et al., 2016). Thus, enhanced VMPFC activity may reflect enhanced sensitivity to information that is valuable to the self in SA, regardless of its valence, and could possibly also indicate excessive scrutiny of one's self-views with regards to that information.

Findings on aberrant VS activity in SA are accumulating as well. Neuroimaging studies of SA found enhanced VS response to social and non-social threats (Boehme et al., 2014; Brühl et al., 2011; Klumpp et al., 2012), as well as VS hypoactivation during anticipation and consumption of social rewards (Becker et al., 2017; Richey et al., 2017, 2014; Schultz et al., 2019). Here we extend this line of findings by revealing altered VS activity in SA during self-evaluation of specific social content, and by suggesting a specific affective process that the VS may mediate during this context. The VS, and specifically the ventral caudate, track the amount of effort people put into gaining rewards (i.e. their degree of 'wanting' rewards) (Croxson, Walton, O'Reilly, Behrens, & Rushworth, 2009; Miller, Shankar, Knutson, & McClure, 2014). Furthermore, both the VS and the VMPFC, which were co-activated here, are central components of a brain system that signals valuable self-relevant information and potentially rewarding outcomes (Bartra et al., 2013; D'Argembeau et al., 2012; Liu et al., 2011). These functions align with the assignment of enhanced motivational significance to power-related content in SA (Kashdan et al., 2008; Roberts et al., 2011). Moreover, high-power traits may constitute a type of 'ideal self' for people with high SA; namely a desirable trait that not only holds high incentive value, but is also uncharacteristic of oneself (Higgins, 1987; Markus & Nurius, 1986). Respectively, people with high SA

may experience a large discrepancy between their perception of their own power and their actual social performance in their everyday lives (Kashdan *et al.*, 2008; Roberts *et al.*, 2011) – a deviance that may also be mediated via the VS and VMPFC (Will, Rutledge, Moutoussis, & Dolan, 2017; Yoon *et al.*, 2018).

Here, the association of VS activity with signaling the high value of high-power traits on the one hand and their mismatch with regards to one's own self-view in SA on the other, were corroborated by two additional analyses. First, stronger VS and VMPFC responses to high-power were also correlated with lower endorsement of these traits. Second, brain activity patterns in the right VS during processing of high-power traits resembled those evoked by incentive salience in the reward task (i.e. a reward anticipation phase), as a function of SA-severity. This may suggest that this type of reward-related processing was activated to some extent in the VS during self-evaluation of high-power as a function of higher SA. However, note that while high-power traits were rated as more positive than low-power traits (see online Supplementary Methods), we did not directly assess participants' subjective evaluations of these traits as desirable or important. Furthermore, it is important to note that previous studies showed that the VS-VMPFC circuit is associated with trait self-esteem (Chavez & Heatherton, 2014, 2017). To exclude the possibility that the correlations we found between LSAS and VS-VMPFC activity are better explained by self-esteem, we show that all results remain significant also while controlling for scores on the self-esteem questionnaire (see online Supplementary Results and Tables S13 and S14).

This study also demonstrates how the methodology of between-task similarity can enhance the interpretability of sub-cortical activity evident during complex cognitive tasks such as self-evaluation. By administering auxiliary fMRI tasks that probe basic affective processes, we were able to demonstrate that activation of similar multi-voxel brain activation patterns across the SRET and the reward task in the VS scaled with SA-severity – thereby suggesting the embedment of a core affective process (i.e. striatal signaling of incentive salience) in self-evaluation. Our approach was inspired by recent studies, which by implementing between-task similarity demonstrated that positive affect and additional mental functions (e.g. language, scene construction) operate during tasks of internal mentation (Axelrod *et al.*, 2017; Chavez *et al.*, 2017). A more recent study showed that participants who activated multivoxel brain activity patterns while viewing neutral pictures that were more similar to those activated during exposure to negative pictures, reported on more positive affect and less negative affect in their daily lives (Puccetti *et al.*, 2021). Together with the results presented here, these findings demonstrate how individual differences in the degree of similarity between a mental state and a certain affective context, may help provide a more precise delineation of the affective neural processes that govern psychopathology symptoms.

Findings from this study may have clinical implications. SA treatment could benefit from identifying problematic self-views underlying this condition. Clinicians may focus on the way patients relate to high-power content, as well as on resolving conflictual motivations associated with such self-views. At the neural level, hitherto studies examining neural changes related to therapeutic effects in SA have largely focused on attenuating amygdala reactivity and on up-regulating fronto-parietal cortical networks that support explicit emotional regulation (Dixon *et al.*, 2020; Goldin *et al.*, 2013; Klumpp *et al.*, 2017; Klumpp & Fitzgerald, 2018; Phan *et al.*, 2013). In contrast, our findings suggest that

downregulating VMPFC and VS activity, possibly mediating the excessive engagement of implicit value-related self processes (e.g. I am not good enough), may serve as an additional clinical target in SA – at least in the context of processing self-related content. Gaining a more nuanced understanding of which neural domain should be targeted under different contexts in order to effectively treat psychopathology symptoms, can also inform novel neuro-modulation based interventions such as process-based neurofeedback (Lubianiker *et al.*, 2019).

Limitations

A number of limitations of this study should be taken into account. The valence of traits in the SRET differed between conditions, as low-intensity conditions consisted mostly of negatively valenced traits whereas high-intensity traits were positively valenced (and were especially positive in the high-affiliation conditions; see online Supplementary Results). Thus, although online Supplementary analyses suggested that traits were categorized and endorsed in a manner that reflected sensitivity to their social domain (see online Supplementary Results and Tables S1 and S2), we cannot determine the extent to which the results were driven by the valence of traits rather than their social domain. Relatedly, the high and uniform endorsement rates of high-affiliation traits suggest that the SRET was less sensitive to measurements of individual differences with respect to this domain. Future studies of power and affiliation could gain better control over these issues by equating valence across categories. In addition, the association of low-intensity traits with their designated category was less robust compared to the high-intensity traits, and it is possible that they could be classified more accurately using a different categorization framework. Future studies could yield more precise profiling of traits by taking into account multiple social domain axes on which they may be represented. Moreover, although clinical levels of SA that meet with the criteria for diagnosing SAD were evident in our sample, it remains an open question whether findings from this study would arise in a SAD patients; and perhaps if such patients would exhibit aberrant activation in regions of interest that were absent here, such as the amygdala.

In conclusion, consistent with cognitive-evolutionary accounts of SA (Gilbert, 2001), our study emphasizes the centrality of power-related self-views in this condition. We show that excessive VS and VMPFC activation accompany self-referential encoding of high-power traits as a function of SA-severity, and suggest that patterns of activations may indicate enhanced significance of such traits for one's ideal self-view. Our findings thus ratify theoretical perspectives that entangle aberrant self-evaluation in psychopathology in general, and in SA in particular, with affective processes (Frewen *et al.*, 2020; Higgins, 1987; Markus & Nurius, 1986).

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291722001453>

Acknowledgements. The authors thank Noa Saron, Shai Igra, Sharon Shilon and Omer Nir for their assistance in conducting the experiment; Dr Noga Cohen for her assistance in constructing the experimental paradigm; Dr Tal Shany-Ur for discussions; and Oren Levin and Shimrit Knirsh (Solnik) for technical support.

Financial support. This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation

under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). The authors would also like to thank the Sagol foundation for funding.

Conflict of interest. None.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

- Aderka, I. M., Weisman, O., Shahar, G., & Gilboa-Schechtman, E. (2009). The roles of the social rank and attachment systems in social anxiety. *Personality and Individual Differences, 47*(4), 284–288. <https://doi.org/10.1016/j.paid.2009.03.014>.
- Axelrod, V., Rees, G., & Bar, M. (2017). The default network and the combination of cognitive processes that mediate self-generated thought. *Nature Human Behaviour, 1*(12), 896–910. <https://doi.org/10.1038/s41562-017-0244-9>.
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427. <https://doi.org/10.1016/j.neuroimage.2013.02.063>.
- Becker, M. P. I., Simon, D., Miltner, W. H. R., & Straube, T. (2017). Altered activation of the ventral striatum under performance-related observation in social anxiety disorder. *Psychological Medicine, 47*(14), 2502–2512. <https://doi.org/10.1017/S0033291717001076>.
- Berger, U., Keshet, H., & Gilboa-Schechtman, E. (2017). Self-evaluations in social anxiety: The combined role of explicit and implicit social-rank. *Personality and Individual Differences, 104*, 368–373. <https://doi.org/10.1016/j.paid.2016.08.023>.
- Blair, K., Geraci, M., Devido, J., McCaffrey, D., Chen, G., Vythilingam, M., ... Pine, D. S. (2008). Neural response to self- and other referential praise and criticism in generalized social phobia. *Archives of General Psychiatry, 65*(10), 1176. <https://doi.org/10.1001/archpsyc.65.10.1176>.
- Blair, K. S., Geraci, M., Otero, M., Majestic, C., Odenheimer, S., Jacobs, M., ... Pine, D. S. (2011). Atypical modulation of medial prefrontal cortex to self-referential comments in generalized social phobia. *Psychiatry Research: Neuroimaging, 193*(1), 38–45.
- Boehme, S., Ritter, V., Tefikow, S., Stangier, U., Strauss, B., Miltner, W. H. R., & Straube, T. (2014). Brain activation during anticipatory anxiety in social anxiety disorder. *Social Cognitive and Affective Neuroscience, 9*(9), 1413–1418. <https://doi.org/10.1093/scan/nst129>.
- Brühl, A. B., Delsignore, A., Komossa, K., & Weidt, S. (2014). Neuroimaging in social anxiety disorder—A meta-analytic review resulting in a new neuro-functional model. *Neuroscience & Biobehavioral Reviews, 47*, 260–280.
- Brühl, A. B., Rufer, M., Delsignore, A., Kaffenberger, T., Jäncke, L., & Herwig, U. (2011). Neural correlates of altered general emotion processing in social anxiety disorder. *Brain Research, 1378*, 72–83. <https://doi.org/10.1016/j.brainres.2010.12.084>.
- Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *NeuroImage, 57*(4), 1608–1616. <https://doi.org/10.1016/j.neuroimage.2011.05.037>.
- Chavez, R. S., & Heatherton, T. F. (2014). Multimodal frontostriatal connectivity underlies individual differences in self-esteem. *Social Cognitive and Affective Neuroscience, 10*(3), 364–370.
- Chavez, R. S., & Heatherton, T. F. (2017). Structural integrity of frontostriatal connections predicts longitudinal changes in self-esteem. *Social Neuroscience, 12*(3), 280–286. <https://doi.org/10.1080/17470919.2016.1164753>.
- Chavez, R. S., Heatherton, T. F., & Wagner, D. D. (2017). Neural population decoding reveals the intrinsic positivity of the self. *Cerebral Cortex, 27*(11), 5222–5229. <https://doi.org/10.1093/cercor/bhw302>.
- Croxxon, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J., & Rushworth, M. F. S. (2009). Effort-based cost-benefit valuation and the human brain. *Journal of Neuroscience, 29*(14), 4531–4541. <https://doi.org/10.1523/JNEUROSCI.4515-08.2009>.
- D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: The valuation hypothesis. *Frontiers in Human Neuroscience, 7*, 372. <https://doi.org/10.3389/fnhum.2013.00372>.
- D'Argembeau, A., Jedidi, H., Baiteau, E., Bahri, M., Phillips, C., & Salmon, E. (2012). Valuing one's self: Medial prefrontal involvement in epistemic and emotive investments in self-views. *Cerebral Cortex, 22*(3), 659–667. <https://doi.org/10.1093/cercor/bhr144>.
- Dawson, J. F., & Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. *Journal of Applied Psychology, 91*(4), 917–926. <https://doi.org/10.1037/0021-9010.91.4.917>.
- Dijk, C., van Emmerik, A. A. P., & Grasman, R. P. P. P. (2018). Social anxiety is related to dominance but not to affiliation as perceived by self and others: A real-life investigation into the psychobiological perspective on social anxiety. *Personality and Individual Differences, 124*, 66–70. <https://doi.org/10.1016/j.paid.2017.11.050>.
- Dixon, M. L., Moodie, C. A., Goldin, P. R., Farb, N., Heimberg, R. G., & Gross, J. J. (2020). Emotion regulation in social anxiety disorder: Reappraisal and acceptance of negative self-beliefs. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5*(1), 119–129. <https://doi.org/10.1016/j.bpsc.2019.07.009>.
- Durnez, J., Degryse, J., Moerkerke, B., Seurinck, R., Sochat, V., Poldrack, R. A., & Nichols, T. E. (2016). Power and sample size calculations for fMRI studies based on the prevalence of active peaks. *bioRxiv*. <https://doi.org/10.1101/049429>.
- Eisenberger, N. I., Inagaki, T. K., Muscatell, K. A., Haltom, K. E. B., & Leary, M. R. (2011). The neural sociometer: Brain mechanisms underlying state self-esteem. *Journal of Cognitive Neuroscience, 23*(11), 3448–3455.
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, 113*(28), 7900–7905. <https://doi.org/10.1073/pnas.1602413113>.
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, 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>.
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., ... Jiang, T. (2016). The human brainnetome atlas: A new brain atlas based on connective architecture. *Cerebral Cortex, 26*(8), 3508–3526. <https://doi.org/10.1093/cercor/bhw157>.
- Fehm, L., Beesdo, K., Jacobi, F., & Fiedler, A. (2008). Social anxiety disorder above and below the diagnostic threshold: Prevalence, comorbidity and impairment in the general population. *Social Psychiatry and Psychiatric Epidemiology, 43*(4), 257–265. <https://doi.org/10.1007/s00127-007-0299-4>.
- Finlayson-Short, L., Harrison, B. J., & Davey, C. (2021). Self-other referential neural processing in social anxiety disorder and major depressive disorder. *NeuroImage: Clinical, 30*, 102669. <https://doi.org/10.1016/j.nicl.2021.102669>.
- Frewen, P., Schroeter, M. L., Riva, G., Cipresso, P., Fairfield, B., Padulo, C., ... Northoff, G. (2020). Neuroimaging the consciousness of self: Review, and conceptual-methodological framework. *Neuroscience & Biobehavioral Reviews, 112*, 164–212. <https://doi.org/10.1016/j.neubiorev.2020.01.023>.
- Frewen, P. A., Lundberg, E., Brimson-Théberge, M., & Théberge, J. (2013). Neuroimaging self-esteem: A fMRI study of individual differences in women. *Social Cognitive and Affective Neuroscience, 8*(5), 546–555. <https://doi.org/10.1093/scan/nss032>.
- Gilbert, P. (2001). Evolution and social anxiety: The role of attraction, social competition, and social hierarchies. *Psychiatric Clinics, 24*(4), 723–751.
- Gilboa-Schechtman, E., Galili, L., Sahar, Y., & Amir, O. (2014). Being 'in' or 'out' of the game: Subjective and acoustic reactions to exclusion and popularity in social anxiety. *Frontiers in Human Neuroscience, 8*, 147.
- Gilboa-Schechtman, E., Keshet, H., Livne, T., Berger, U., Zabag, R., Hermesh, H., & Marom, S. (2017). Explicit and implicit self-evaluations in social anxiety disorder. *Journal of Abnormal Psychology, 126*(3), 285–290. <https://doi.org/10.1037/abn0000261>.
- Gilboa-Schechtman, E., Keshet, H., Peschard, V., & Azoulay, R. (2020). Self and identity in social anxiety disorder. *Journal of Personality, 88*(1), 106–121. <https://doi.org/10.1111/jopy.12455>.

- Goldin, P., Ramel, W., & Gross, J. (2009a). Mindfulness meditation training and self-referential processing in social anxiety disorder: Behavioral and neural effects. *Journal of Cognitive Psychotherapy*, 23(3), 242–257. <https://doi.org/10.1891/0889-8391.23.3.242>.
- Goldin, P., Ziv, M., Jazaieri, H., & Gross, J. (2012). Randomized controlled trial of mindfulness-based stress reduction versus aerobic exercise: Effects on the self-referential brain network in social anxiety disorder. *Frontiers in Human Neuroscience*, 6, 295. <https://doi.org/10.3389/fnhum.2012.00295>.
- Goldin, P. R., & Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion (Washington, D.C.)*, 10(1), 83–91. <https://doi.org/10.1037/a0018441>.
- Goldin, P. R., Manber-Ball, T., Werner, K., Heimberg, R., & Gross, J. J. (2009b). Neural mechanisms of cognitive reappraisal of negative self-beliefs in social anxiety disorder. *Biological Psychiatry*, 66(12), 1091–1099.
- Goldin, P. R., Ziv, M., Jazaieri, H., Hahn, K., Heimberg, R., & Gross, J. J. (2013). Impact of cognitive behavioral therapy for social anxiety disorder on the neural dynamics of cognitive reappraisal of negative self-beliefs: Randomized clinical trial. *JAMA Psychiatry*, 70(10), 1048–1056. <https://doi.org/10.1001/jamapsychiatry.2013.234>.
- Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: A comparison of faces and scenes. *NeuroImage*, 17(1), 317–323.
- Heitmann, C. Y., Peterburs, J., Mothes-Lasch, M., Hallfarth, M. C., Böhme, S., Miltner, W. H., & Straube, T. (2014). Neural correlates of anticipation and processing of performance feedback in social anxiety. *Human Brain Mapping*, 35(12), 6023–6031.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319–340. <https://doi.org/10.1037/0033-295X.94.3.319>.
- Hughes, B. L., & Beer, J. S. (2013). Protecting the self: The effect of social-evaluative threat on neural representations of self. *Journal of Cognitive Neuroscience*, 25(4), 613–622.
- Hyett, M. P., & McEvoy, P. M. (2018). Social anxiety disorder: Looking back and moving forward. *Psychological Medicine*, 48(12), 1937–1944. <https://doi.org/10.1017/S0033291717003816>.
- Izuma, K., Kennedy, K., Fitzjohn, A., Sedikides, C., & Shibata, K. (2018). Neural activity in the reward-related brain regions predicts implicit self-esteem: A novel validity test of psychological measures using neuroimaging. *Journal of Personality and Social Psychology*, 114(3), 343–357. <https://doi.org/10.1037/pspa0000114>.
- Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience*, 5(2), 133–147.
- Kashdan, T. B., Elhai, J. D., & Breen, W. E. (2008). Social anxiety and disinhibition: An analysis of curiosity and social rank appraisals, approach-avoidance conflicts, and disruptive risk-taking behavior. *Journal of Anxiety Disorders*, 22(6), 925–939. <https://doi.org/10.1016/j.janxdis.2007.09.009>.
- Klumpp, H., Angstadt, M., & Phan, K. L. (2012). Insula reactivity and connectivity to anterior cingulate cortex when processing threat in generalized social anxiety disorder. *Biological Psychology*, 89(1), 273–276. <https://doi.org/10.1016/j.biopsycho.2011.10.010>.
- Klumpp, H., & Fitzgerald, J. M. (2018). Neuroimaging predictors and mechanisms of treatment response in social anxiety disorder: An overview of the amygdala. *Current Psychiatry Reports*, 20(10), 1–9. <https://doi.org/10.1007/s11920-018-0948-1>.
- Klumpp, H., Fitzgerald, J. M., Kinney, K. L., Kennedy, A. E., Shankman, S. A., Langenecker, S. A., & Phan, K. L. (2017). Predicting cognitive behavioral therapy response in social anxiety disorder with anterior cingulate cortex and amygdala during emotion regulation. *NeuroImage: Clinical*, 15, 25–34. <https://doi.org/10.1016/j.nicl.2017.04.006>.
- Leichsenring, F., & Leweke, F. (2017). Social anxiety disorder. *New England Journal of Medicine*, 376(23), 2255–2264. <https://doi.org/10.1056/NEJMc1614701>.
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, 99, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>.
- Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, 22, 141–173.
- Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 35(5), 1219–1236.
- Lubianiker, N., Goldway, N., Fruchtman-Steinbok, T., Paret, C., Keynan, J. N., Singer, N., ... Hendler, T. (2019). Process-based framework for precise neuromodulation. *Nature Human Behaviour*, 3(5), 436–445. <https://doi.org/10.1038/s41562-019-0573-y>.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954.
- Mennin, D. S., Fresco, D. M., Heimberg, R. G., Schneier, F. R., Davies, S. O., & Liebowitz, M. R. (2002). Screening for social anxiety disorder in the clinical setting: Using the Liebowitz social anxiety scale. *Journal of Anxiety Disorders*, 16(6), 661–673. [https://doi.org/10.1016/S0887-6185\(02\)00134-2](https://doi.org/10.1016/S0887-6185(02)00134-2).
- Miller, E. M., Shankar, M. U., Knutson, B., & McClure, S. M. (2014). Dissociating motivation from reward in human striatal activity. *Journal of Cognitive Neuroscience*, 26(5), 1075–1084.
- Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of Cognitive Neuroscience*, 18(9), 1586–1594.
- Moscovitch, D. A. (2009). What is the core fear in social phobia? A new model to facilitate individualized case conceptualization and treatment. *Cognitive and Behavioral Practice*, 16(2), 123–134. <https://doi.org/10.1016/j.cbpra.2008.04.002>.
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, 10(4), e1003553. <https://doi.org/10.1371/journal.pcbi.1003553>.
- Norton, A. R., & Abbott, M. J. (2016). Self-focused cognition in social anxiety: A review of the theoretical and empirical literature. *Behaviour Change*, 33(1), 44–64. <https://doi.org/10.1017/bec.2016.2>.
- Peterburs, J., Sandrock, C., Miltner, W. H., & Straube, T. (2016). Look who's judging – feedback source modulates brain activation to performance feedback in social anxiety. *NeuroImage*, 133, 430–437.
- Phan, K. L., Coccaro, E. F., Angstadt, M., Kreger, K. J., Mayberg, H. S., Liberzon, I., & Stein, M. B. (2013). Corticolimbic brain reactivity to social signals of threat before and after sertraline treatment in generalized social phobia. *Biological Psychiatry*, 73(4), 329–336. <https://doi.org/10.1016/j.biopsych.2012.10.003>.
- Puccetti, N. A., Schaefer, S. M., van Reekum, C. M., Ong, A. D., Almeida, D. M., Ryff, C. D., ... Heller, A. S. (2021). Linking amygdala persistence to real-world emotional experience and psychological well-being. *Journal of Neuroscience*, 41(16), 3721–3730. <https://doi.org/10.1523/JNEUROSCI.1637-20.2021>.
- Putnam, P. T., & Gothard, K. M. (2019). Multidimensional neural selectivity in the primate amygdala. *Eneuro*, 6(5), 1–13. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6791828/pdf/ENEURO.0153-19.2019.pdf>.
- Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, 57(3), 1221–1233. <https://doi.org/10.1016/j.neuroimage.2011.05.028>.
- Richey, J. A., Ghane, M., Valdespino, A., Coffman, M. C., Stregue, M. V., White, S. W., & Ollendick, T. H. (2017). Spatiotemporal dissociation of brain activity underlying threat and reward in social anxiety disorder. *Social Cognitive and Affective Neuroscience*, 12(1), 81–94. <https://doi.org/10.1093/scan/nsw149>.
- Richey, J. A., Rittenberg, A., Hughes, L., Damiano, C. R., Sabatino, A., Miller, S., ... Dichter, G. S. (2014). Common and distinct neural features of social and non-social reward processing in autism and social anxiety disorder. *Social Cognitive and Affective Neuroscience*, 9(3), 367–377. <https://doi.org/10.1093/scan/nss146>.
- Roberts, K. E., Hart, T. A., Coroiu, A., & Heimberg, R. G. (2011). Gender role traits among individuals with social anxiety disorder. *Personality and Individual Differences*, 51(8), 952–957. <https://doi.org/10.1016/j.paid.2011.07.030>.
- Ruscio, A. M. (2010). The latent structure of social anxiety disorder: Consequences of shifting to a dimensional diagnosis. *Journal of Abnormal Psychology*, 119(4), 662–671. <https://doi.org/10.1037/a0019341>.
- Rytwinski, N. K., Fresco, D. M., Heimberg, R. G., Coles, M. E., Liebowitz, M. R., Cissell, S., ... Hofmann, S. G. (2009). Screening for social anxiety

- disorder with the self-report version of the Liebowitz social anxiety scale. *Depression and Anxiety*, 26(1), 34–38. <https://doi.org/10.1002/da.20503>.
- Schultz, J., Willems, T., Gädeke, M., Chakkour, G., Franke, A., Weber, B., ... Hurlemann, R. (2019). A human subcortical network underlying social avoidance revealed by risky economic choices. *eLife*, 8, e45249. <https://doi.org/10.7554/eLife.45249>.
- Scult, M. A., Knodt, A. R., Radtke, S. R., Brigidi, B. D., & Hariri, A. R. (2019). Prefrontal executive control rescues risk for anxiety associated with high threat and low reward brain function. *Cerebral Cortex*, 29(1), 70–76. <https://doi.org/10.1093/cercor/bhx304>.
- Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, 1(1), 1–10. <https://doi.org/10.1038/s42003-018-0073-z>.
- Weisman, O., Aderka, I. M., Marom, S., Hermesh, H., & Gilboa-Schechtman, E. (2011). Social rank and affiliation in social anxiety disorder. *Behaviour Research and Therapy*, 49(6–7), 399–405. <https://doi.org/10.1016/j.brat.2011.03.010>.
- Will, G.-J., Moutoussis, M., Womack, P. M., Bullmore, E. T., Goodyer, I. M., Fonagy, P., ... Dolan, R. J. (2020). Neurocomputational mechanisms underpinning aberrant social learning in young adults with low self-esteem. *Translational Psychiatry*, 10(1), 1–14. <https://doi.org/10.1038/s41398-020-0702-4>.
- Will, G.-J., Rutledge, R. B., Moutoussis, M., & Dolan, R. J. (2017). Neural and computational processes underlying dynamic changes in self-esteem. *eLife*, 6:e28098. <https://doi.org/10.7554/eLife.28098>.
- Wittchen, H. U., Fuetsch, M., Sonntag, H., Müller, N., & Liebowitz, M. (2000). Disability and quality of life in pure and comorbid social phobia. Findings from a controlled study. *European Psychiatry*, 15(1), 46–58. [https://doi.org/10.1016/S0924-9338\(00\)00211-X](https://doi.org/10.1016/S0924-9338(00)00211-X).
- Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419. <https://doi.org/10.1016/j.neuroimage.2013.12.058>.
- Yoon, H.-J., Seo, E. H., Kim, J.-J., & Choo, I. H. (2019). Neural correlates of self-referential processing and their clinical implications in social anxiety disorder. *Clinical Psychopharmacology and Neuroscience*, 17(1), 12–24. <https://doi.org/10.9758/cpn.2019.17.1.12>.
- Yoon, L., Somerville, L. H., & Kim, H. (2018). Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nature Communications*, 9(1), 3086. <https://doi.org/10.1038/s41467-018-05553-2>.