Abstract: 196
Number of Words: 4733
Number of Figures: 3
Number of Tables: 2

**Neural differences in self-perception during illness and after weight-recovery in anorexia nervosa**

**Authors:** Carrie J. McAdams M.D. Ph.D. 1,2*, Haekyung Jeon-Slaughter, Ph.D.1, Siobahn Evans1, Terry Lohrenz, Ph.D.3, P. Read Montague Ph.D.3,4,5,6, Daniel C. Krawczyk Ph. D.1,7

1 Department of Psychiatry, University of Texas at Southwestern Medical School, Dallas, TX 75390, USA
2 Psychiatry, Texas Health Presbyterian Hospital of Dallas, Dallas, TX, 75321, USA
3 Virginia Tech Carilion Research Institute, Virginia Tech, Roanoke, VA 24016, USA
4 Computational Psychiatry Unit, Virginia Tech, Roanoke, VA 24016, USA
5 Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, WCIN 3BG, London, UK
6 Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA
7 Center for Brain Health, University of Texas at Dallas, Dallas, TX 75390, USA

*Corresponding:

Carrie J. McAdams, MD PhD
Department of Psychiatry
University of Texas at Southwestern Medical School

© The Author (2016). Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Acknowledgments: Funding for this research was provided by K23 MH093684 and a 2012 NARSAD Young Investigator Award to Dr. McAdams. We thank all subjects that have participated, as well as Dr. Sunbola Ashimi, Jarrette Moore, and Lauren Monier for their work as research coordinators. We appreciate comments on the manuscript from Dr. Carol Tamminga. The authors have no conflicts of interest to disclose.

Running Title: Self-perception in anorexia nervosa
ABSTRACT

Anorexia nervosa is a severe mental illness characterized by problems with self-perception. Whole-brain neural activations in healthy women, women with anorexia nervosa, and women in long-term weight recovery following anorexia nervosa were compared using two functional magnetic resonance imaging tasks probing different aspects of self-perception. The Social Identity-V2 task involved consideration about oneself and others using socially descriptive adjectives. Both the ill and weight-recovered women with anorexia nervosa engaged medial prefrontal cortex less than healthy women for self-relevant cognitions, a potential biological trait difference. Weight-recovered women also activated the inferior frontal gyri and dorsal anterior cingulate more for direct self-evaluations than for reflected self-evaluations, unlike both other groups, suggesting that recovery may include compensatory neural changes related to social perspectives. The Faces task compared viewing oneself to a stranger. Participants with anorexia nervosa showed elevated activity in the bilateral fusiform gyri for self-images, unlike the weight-recovered and healthy women, suggesting cognitive distortions about physical appearance are a state rather than trait problem in this disease. Because both ill and recovered women showed neural differences related to social self-perception, but only recovered women differed when considering social perspectives, these neurocognitive targets may be particularly important for treatment.
Keywords: Eating Disorders, Medial Prefrontal Cortex, self-reflection, fMRI, psychiatry
INTRODUCTION

Anorexia nervosa (AN) is a serious psychiatric illness associated both with an inability to maintain one’s body weight, and altered perceptions about one’s body shape and weight (American Psychiatric Association, 1994). Problems in the internal construct of self, including both social and physical aspects of one’s identity, have been proposed to be a core feature of this illness leading to food restriction (Fairchild & Cooper, 2010; Stein & Corte, 2007). A failure to develop a secure sense of identity might promote disordered eating behaviors in many ways.

Acceptance of one’s actual body shape and size, a difficult problem in AN (Skrzypek, Wehmeier, & Remschmidt, 2001), requires an internal sense of physical self. Identifying why one’s emotions might be distressing requires understanding of one’s emotions. However, starvation appears to reduce distressing emotions in AN, a factor that may exacerbate the disease (Brockmeyer et al., 2011), and alexithymia, the ability to identify emotions, is elevated in AN (Beadle, Paradiso, Salerno, & McCormick, 2013). Women with AN also show high needs for external sources of approval (Amianto, Abbate-Daga, Morando, Sobrero, & Fassino, 2011); that approval might be sought by seeking out socially-imposed beliefs such as the thin ideal.

Problems in making decisions (Guillaume et al., 2015), and navigating interpersonal relationships (Treasure & Schmidt, 2013) are also observed in AN; both behaviors require self-reflection to identify how pros and cons of a situation relate to oneself. Fundamentally, decisions related to self – emotions and values - are ubiquitous, and might contribute to many characteristics and traits that lead to both the restricted food intake and preoccupations about body image that characterize AN.

Currently, there is a limited understanding of the pathology of AN, and even less understanding of how pathology and treatments are related (Dalle Grave, 2011). Although
weight-restoration is essential for recovery from AN, more than half of adult patients relapse after treatment (Keel & Brown, 2010; Steinhausen, 2002). Therapies that consider body-image, identity, and self-esteem are typically utilized after weight-restoration but the specific cognitive, physiological, and emotional changes that result in sustained recovery have not been clearly identified (Brockmeyer et al., 2013; Hay et al., 2014; Kelly, Vimalakanthan, & Carter, 2014; Lazaro et al., 2011; Murphy, Straebler, Cooper, & Fairburn, 2010; Watson & Bulik, 2013). Utilization of functional magnetic resonance imaging (fMRI) tasks to characterize specific neurocognitive problems in self-perception in AN may improve our ability to understand the disease and design targeted treatments.

Neurobiologically, medial prefrontal cortex, posterior cingulate cortex, and the temporoparietal junctions have been closely related to self-perception and evaluation in healthy populations (Denny, Kober, Wager, & Ochsner, 2012). Differences in neural activations during social and self-perception in these regions have been reported in AN (McAdams & Krawczyk, 2011, 2014; McCormick et al., 2008; McFadden, Tregellas, Shott, & Frank, 2014; Miyake et al., 2012; Miyake et al., 2010; Muhlau et al., 2007; Schulte-Ruther, Mainz, Fink, Herpertz-Dahlmann, & Konrad, 2012). In the present study, we were interested in determining whether women with and recovered from AN showed similar neural activations during social and physical self-perception. The study goal was to assess whether neural differences related to self-perception might provide a better understanding of the pathogenesis of the disease and recovery from AN.

We compared neural activations to self-stimuli in three groups of subjects: healthy comparison women, women with AN and women in long-term weight-recovery following AN. Using these groups, we evaluated whether neural differences in processing self-stimuli could be
associated with either the development of the disease (present in both ill and recovered but not healthy women) or stage of the disease (present in ill or recovered but not both patient groups).

We adapted the social appraisal task from our prior study (McAdams & Krawczyk, 2014), hypothesizing that women with and recovered from AN would show frontal lobe differences during social self-evaluations. We utilized a self and stranger face-viewing task to probe neural activity during physical self-reflection (Kishida, Li, Schwind, & Montague, 2012), hypothesizing that self-image would more strongly engage visual regions during the illness than after recovery.

**METHODS**

**Participants**

Subjects provided written informed consent to participate, approved by the UT Southwestern Institutional Review Board. A total of 59 women (age range 20-45 years) participated: 19 healthy comparison women (HC), 22 women currently with AN (AN-C, DSM IV criteria within prior 6 months), and 18 women with sustained weight recovery (AN-WR, 24 months with BMI > 19.0). Subjects were interviewed using the Structured Clinical Interview for DSM-IV to confirm course and history of anorexia nervosa for AN-C and AN-WR groups, and absence of eating disorders (ED) in the HC group. More clinical details about the groups are provided in the supplement (Methods S1).

Subjects completed a structured interview for current depression (Quick Inventory of Depressive Symptoms, QIDS-CR), and anxiety (Structured Interview Guide for the Hamilton Anxiety Scale, SIGH-A). Eating disorder symptoms were assessed with the Eating Disorder Examination Questionnaire (EDE-Q) in the AN-C and AN-WR group; all subjects also
completed the Eating Attitudes Test (EAT). The Wechsler Abbreviated Scale of Intelligence (WASI) provided an estimate of intelligence.

Neuroimaging Tasks

The Social Identity-V2 task consisted of three different conditions: Self (ex. “I believe I am deceitful”), Friend (“I believe my friend is moody”), and Reflected (“My friend believes I am proud”), personalized with the name of a female friend. Each statement was presented above the terms Agree and Disagree for 4 s, and a response was selected each trial, followed by a jittered fixation period of 4, 6, or 8 s. There were 48 statements in each of three runs; 16 statements for each condition resulting in a total duration of 8 minutes per run. All statements for each condition were sequential, and condition order pseudorandomized across runs. The Social Identity-V2 task differed from the original Social Identity task (McAdams & Krawczyk, 2014) (See Methods S2). During each trial, each subject indicated a rating of agreement using a hand-held button, yielding behavioral data as a reaction time and response (Agree or Disagree).

For the Faces task, each subject viewed fifteen images of their own face (F-Self) and fifteen images of a stranger (F-Other), each presented for 4 s with a jittered time-interval (Kishida et al., 2012). The stranger image matched the subject as closely as possible in age, skin-tone, and hair color; all images differed but included both eyes open, a neutral or positive expression, with a variety of head tilts. One image appeared every 8 to 30 s, with a fixation cross between images. Because of technical difficulties, four women in the AN-C group did not complete the Faces task.
Statistical Analyses of Clinical and Behavioral Data

Demographic, clinical and behavioral comparisons were completed in SPSS v.21 (SPSS, Inc., Chicago). Measures obtained for all groups (Age, BMI, WASI, QIDS, SIGH-A, EAT) were compared using analysis of variance (ANOVA) with post-hoc comparisons, using a Bonferroni adjusted p < 0.05. Measures collected only from the patient groups (EDE-Q, age of onset) were compared with a two-sample t-test (p < 0.05). Mean reaction times and mean percentage of trials showing agreement in the Social Identity-V2 task were compared across groups and conditions using multivariate analysis of variance (ANOVA) followed by Bonferroni post-hoc comparisons (criterion of p < 0.05).

MRI Acquisition and Analysis

Images were acquired with a 3T Philips Achieva MRI scanner, using a 1-shot gradient T2*-weighted echoplanar (EPI) image sequence sensitive to blood oxygen level-dependent (BOLD) contrast with a repetition time (TR) of 2 s. For Social Identity-V2, the echo time (TE) was 35 ms, and the flip angle was 0°; volumes were composed of 36 axial slices (4 mm thick, no gap). For Faces, the echo time (TE) was 30 ms, and a flip angle was 90°; volumes were composed of 34 axial slices (4 mm thick, no gap). For both tasks, each slice was acquired with a 22.0 cm² field of view, a matrix size of 64x64 and a voxel size of 3.4x3.4x4 mm. Functional images were acquired during 4 runs (3 for Social Identity-V2, each 480 s; 1 for Faces, between 540 and 600 s). High resolution MP-RAGE 3D T1-weighted images were acquired for anatomical localization with the following imaging parameters: TR = 2100 ms, TE = 3.7 ms; slice thickness of 1 mm with no gap, a 12° flip angle, and 1 mm³ voxels.
Preprocessing consisted of spatial realignment to the first volume of acquisition, normalization to the MNI standard template, and spatial smoothing with a 6 mm 3D Gaussian kernel. The voxel time-series was high pass filtered (128 s). Functional MRI task data were analyzed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Imaging Neuroscience London, www.fil.ion.ucl.ac.uk/spm) run in Matlab 2012 (http://www.mathworks.com) and viewed in both SPM and xjview (http://www.alivelearn.net/xjview8/). An event-related design extracted the BOLD signal during the 4-s presentation of each statement or image and a general linear model created contrast images of each event (events: Social Identity-V2: Self-Agree, Self-Disagree, Friend-Agree, Friend-Disagree, Reflected-Agree, Reflected-Disagree; Faces: F-Self, F-Other). Evoked activation was assessed using multiple regression analysis set as boxcar functions. Each regressor was convolved with a canonical hemodynamic response function (HRF) provided in SPM8 and entered into the modified general linear model (GLM). Parameter estimates (e.g. β values) were extracted from this GLM analysis for each regressor. A first level analysis evaluated task contrasts (Social Identity-V2, three contrasts: Self – Friend, Self-Agree – Self-Disagree, Reflected – Self; Faces task, one contrast: F-Self – F-Other). Supplementary analyses were also completed for the Social Identity-V2 task, considering valence of the adjectives rather than behavioral responses (Supplemental Analysis S1).

To identify group differences, a whole-brain ANOVA followed by pairwise comparisons identified regions of interest (ROI) (threshold: cluster-corrected \( p_{\text{FWE}} < 0.05 \) with voxel-wise \( p < 0.005 \), uncorrected). Age was a covariate of no interest for the whole-brain analyses. For each ROI showing significant group differences from the exploratory whole-brain comparisons, parameter estimates (β values) were extracted for each subject using Marsbar (Brett, Anton,
Valabregue, & Poline, 2002). The parameter estimates were exported to SPSS, and the groups compared using an ANOVA followed by pairwise comparisons (Bonferroni-corrected p < 0.05).

Two methods considered whether clinical measures were related to the neural activations in the tasks. Pearson’s correlations considered whether clinical symptoms (QIDS-CR, SIGH-A, EDE-Q) correlated with the extracted parameter estimates within the seven ROIs obtained from the whole-brain ANOVAs. These correlations were performed across all patient-participants (AN-C and AN-WR), and separately for each patient group (corrected threshold: p value of 0.007 (p = 0.05 / 7 ROIs)). Second, whole-brain regressions explored whether the first-level images for each of the four contrasts correlated with the clinical symptoms for the patient-participants only (cluster P$_{FWE}$ < 0.05, with voxel-wise P < 0.005). Psychoactive medication effects were considered with a group x medication multivariate ANOVA using extracted parameter estimates for patient-subjects (criterion: p value of 0.007 (p = 0.05 / 7 ROIs)).

RESULTS

Demographic and Clinical Comparisons.

Clinical and demographic differences are provided in Table 1. The AN-C group had a lower body mass index than both the HC and AN-WR groups. All three groups differed on the overall EAT, and its Dieting and Bulimia Subscales. The AN-C and AN-WR groups differed from the HC group, but not each other, for the EAT Oral Subscale, QIDS-CR and SIGH-A. The AN-C group had more symptoms on the EDE-Q and all subscales compared to the AN-WR group. In sum, all significant psychiatric differences in the groups were related to eating disorder symptoms, not anxiety or depression symptoms.
The Social Identity-V2 Task

**Behavioral Data.** Response (*Agree* or *Disagree*) and reaction time for each condition (*Self*, *Friend*, and *Reflected*) were obtained each trial (Table S1). There were no group or condition differences in reaction times, indicating that all groups spent similar amounts of time evaluating each statement. The AN-C group showed less agreement across all conditions compared to both the AN-WR and HC groups (percent agree, AN-C = 52%; AN-WR = 56%, HC = 55%; F(2,56) = 7.094, p = 0.001). All groups agreed more during *Self* than for *Friend* and *Reflected* (percent agree, Self = 57%, Friend = 52%, Reflected = 53%; F(2,56) = 11.847, p < 0.001).

**Social Self-Perception: Self – Friend.** The *Self – Friend* contrast directly compares activation of neural regions for self-evaluations with friend-evaluations. Across all groups, direct evaluation of oneself activated regions in the bilateral parietal cortex, whereas evaluations of one’s friend activated medial prefrontal cortex and the posterior cingulate/precuneus (Figure S1). Considering each group separately, the HC group had more clusters for *Friend* than *Self*; the AN-C group showed no neural differences; and the AN-WR group showed more activations for *Self* than *Friend* (Table S2). A region in the medial prefrontal cortex extending into the dorsal anterior cingulate (MPFC-dACC) differed in the whole-brain ANOVA comparing all groups (Figure 1A, Table 2). This area was more active during *Self* trials for the AN-WR group and more active during *Friend* trials for the HC group, with little modulation in the AN-C group. Similarly in pairwise comparisons, both the AN-WR – HC and AN-C – HC comparisons had more neural activity for *Self* than *Friend* and no differences were observed in the AN-C and AN-WR comparisons (Table S3).
Social Self-Relevance: Self-Agree – Self-Disagree. The Self-Agree – Self-Disagree contrast isolates neural activations related to self-relevance of the stimuli (Table S4). Group differences in the MPFC extending into the cingulate were observed, with the extracted parameter estimates showing that the HC utilized this area more for Self-Agree trials but both the AN-C and AN-WR activated it more for Self-Disagree trials (Table 2, Figure 1B; Table S5). These results suggest that women with AN, both during illness and after sustained recovery, use MPFC-Cing more for cognitive processes related to self-disagreement whereas healthy women use this area more for cognitive processes related to self-agreement.

Social Evaluation: Reflected – Self. The Reflected – Self contrast typically leads to the recruitment of additional social cognitive neural regions (D'Argembeau et al., 2007), and this was observed in the all subjects group map (Table S6 and Figure S2). In the group comparison, the bilateral inferior frontal gyri extending into the anterior insulae (LIFG, RIFG), and the dorsal anterior cingulate (dACC) differed for all three groups (Table 2, Figure 2). Consideration of the β values from these regions shows that the AN-WR group engaged all regions more for Self than for Reflected evaluations (negative values in Figure 2B for Reflected-Self contrast), and differed from the HC and AN-C group in all regions. The AN-C group overall showed relatively little differentiation in the Reflected – Self contrast (values near zero), and differed from the HC group only in the LIFG. Pairwise comparisons supported these results with differences in the HC- AN-C contrast, the HC – AN-WR contrast, and in the AN-C – AN-WR contrasts (Table S7).

The Faces Task
In the *Faces* task, subjects passively viewed their own image (*F-Self*) or a comparable picture of a stranger (*F-Other*). Across all subjects and in the AN-WR and HC groups, presentation of one’s own image led to stronger activations of the middle cingulate, bilateral insula and inferior frontal gyri whereas viewing the stranger’s image activated the bilateral temporoparietal junctions and precuneus (Table S8 and Figure S3). The AN-C group did not show any regions with elevated activity for the *F-Other* trials, and instead had clusters in the temporal and parietal regions in the *F-Self* trials. Whole-brain differences were observed only in the pairwise comparison of the AN-C and AN-WR groups (Table 2, Figure 3). One cluster was in the left middle temporal gyrus, extending into the fusiform gyrus (LFusi), and the other was in the right middle temporal gyrus, extending into the fusiform gyrus (RFusi). In both cases, the extracted parameter estimates from these ROIs confirmed more F-Self activity in the AN-C group than in both the HC and AN-WR groups (Figure 3).

**Clinical Symptoms, Medications, and Neural Activations**

Parameter estimates were extracted from the seven ROIs identified in the whole-brain group comparisons and correlated with the clinical symptom scales (EDE-Q, SIGH-A, QIDS). Consistent with the whole-brain results, several of the regions that differed for the AN-C and AN-WR groups were also correlated with the EDE-Q-Global (RIFG, \( r = 0.488, p = 0.001 \); dACC, \( r = 0.441, p = 0.004 \); LFusi, \( r = 0.587, p < 0.001 \); RFusi, \( r = 0.474, p = 0.003 \)). However, this relationship was not present within any region when each patient group was considered separately. Neither the depression (QIDS) nor anxiety (SIGH-A) scales correlated with any ROIs. Second, because there were four contrasts and three clinical scales, twelve whole-brain
regressions explored whether clinical scales directly correlated with whole-brain responses. Three clusters were identified in the social self-evaluation contrast (Social Identity-V2, Reflected – Self) regression with the EDE-Q global scale (Table 2, Supplemental Figure 4). No clusters were identified in any of the regressions for the other three contrasts or related to the depression and anxiety scales. The group x medication use ANOVA found no significant differences related to medication use (Table S9).

DISCUSSION

We examined how the neural processes engaged during four cognitive contrasts related to self-perception differed amongst women with AN, those with sustained weight-recovery from AN, and healthy women (See Patient Perspectives, S1). First, in the self-relevance contrast, both ill and weight-recovered women utilized medial prefrontal cortex and the cingulate more when disagreeing about descriptors whereas the HC group utilized these regions when agreeing with the terms. In this area and contrast alone, the AN-C and AN-WR groups were similar to each other but differed consistently from the HC group, suggesting a persistent biological trait difference related to self-relevance. Second, the weight-recovered women showed different neural activations compared to both the ill and the healthy women in several frontal regions during the two conditions related to social self-perception and evaluation. We hypothesized that both groups with AN would show altered engagement of frontal regions during social self-evaluation. That hypothesis was formulated because alterations in the salience network including both the dorsal anterior cingulate and bilateral insula have been observed in many psychiatric illnesses (Downar, Blumberger, & Daskalakis, 2016; Goodkind et al., 2015). Instead, we found differences in the utilization of these salience network regions during social self-evaluation only
for the AN-WR group, suggesting a neural difference that may be important for sustained weight-recovery. Finally, we hypothesized that ill women might show more neural activations in visual regions when viewing themselves compared to others, relative to the other groups, because problems in physical self-perception are associated with acute illness. We confirmed that hypothesis, observing increased bilateral fusiform activations for self-stimuli in the ill but not the other groups.

MPFC was utilized for different cognitive functions by the HC group (more active in Friend and Self-Agree trials) than by both the AN-C and AN-WR groups (more active in Self and Self-Disagree trials). This result suggests that the biological processes that mediate social self-perception and self-relevance may be fundamentally different in AN, indicative of a possible trait related to development of the disease. van Veluw and Chance (van Veluw & Chance, 2014) conducted a meta-analysis proposing that MPFC and the superior temporal gyri are key neural regions involved in self-recognition. Oberndorfer et al. (Oberndorfer, Kaye, Simmons, Strigo, & Matthews, 2011) reported reduced ability to modulate MPFC in response to task difficulty amongst weight-recovered women with AN, supporting the idea that dysfunction of MPFC may contribute to development of this disease. Another neural region that has been proposed as a trait related to pathology in AN is the ventral striatum (Frank, 2015). Intriguingly, in healthy men and women, the connectivity of MPFC to the ventral striatum has been associated with trait self-esteem, whereas functional activation of the MPFC for self-relevance correlated with state self-esteem (Chavez & Heatherton, 2015). Self-esteem has been related to eating disorder severity, social problem-solving, and outcomes (Karpowicz, Skarsater, & Nevonen, 2009; Paterson, Power, Yellowlees, Park, & Taylor, 2007). In concert, these data suggest that dysfunction in corticostriatal circuits contribute to both symptoms of disordered eating as well as the alterations
in self-perception observed in AN. Without a strong sense of reward when experiencing either positive social or physiological (food) feedback, women with AN may both lack both a secure sense of self-identity and motivation to eat. Future work to connect self-esteem, corticostriatal circuitry, and illness trajectory in AN may lead to a mechanistic model of how neurobiological factors contribute to the development of AN.

One of the study goals was to determine if regional neural responses related to self-stimuli differed for the AN-C and AN-WR groups because understanding how changes related to self-perception are related to clinical course may help with treatment recommendations and targets. These differences could result in normalization of a brain region that was previously dysfunctional during illness (AN-WR and HC same, AN-C differs) or changes in brain regions only amongst the sustained weight-recovered group (AN-WR differs, HC and AN-C same). Both types of results were observed in this study.

In the physical self-perception task, we found that the AN-C group had increased processing of self-images compared to stranger-images in visual regions relative to the HC and AN-WR groups, suggesting that sustained-weight recovery normalizes neural responses for physical self-perception. Consistent with these results, Inoue et al. (Inoue et al., 2015) utilized near-infrared spectroscopy to examine right temporal lobe responses to self and stranger faces in adolescents with AN, finding increased blood flow to these regions for face stimuli. In sum, these data suggest both that hyperactivation of early visual regions for self-stimuli may alter physical self-perception during acute disease and that biological aspects of physical self-perception may be resolved after sustained weight-recovery. This is encouraging, as it suggests physical distortions related to body-image are transient rather than persistent traits that
predispose to AN, or that treatments currently utilized for AN succeed, resolving the body-image components in this disease.

Alternatively, in the social self-evaluation contrast of the Social Identity-V2 task, we observed differences in activations in three frontal regions, the bilateral anterior insula and dorsal anterior cingulate, for the AN-WR group relative to both other groups. The dorsal anterior cingulate and anterior insula are both parts of the salience network, and have been proposed to serve as a switch related to motivation and environment that is disrupted in mental illness (Downar et al., 2016). Changes in utilization of the salience network for self-stimuli may be important for sustained weight-recovery following AN. Importantly, there were no behavioral differences for the Social Identity-V2 task in the AN-WR and HC comparisons. This suggests that neural activations can be more sensitive in detecting internal cognitive differences than behavioral measures.

Because this was a cross-sectional rather than longitudinal study, differences in the AN-WR group relative to the other groups could reflect neural compensations, pre-existing differences predictive of recovery, or a biological trait suppressed during acute disease. We favor the neural compensation hypothesis because outpatient treatment of AN typically includes individual psychotherapy and cognitive therapies that focus on self-perception, body-image, mentalization, and self-esteem (Brockmeyer et al., 2013; Hay et al., 2014; Kelly et al., 2014; Lazaro et al., 2011; Watson & Bulik, 2013), all constructs related to social self-evaluations. One area that differed during reflected 3rd person evaluations compared to direct 1st person evaluations in the recovered women was the dorsal anterior cingulate, a region that has been directly related to the difficulty of a task (Shenhav, Straccia, Cohen, & Botvinick, 2014). This suggests that after weight-recovery from AN, self-evaluations may be more challenging, perhaps
due to internal conflicts in self-assessments. A recent meta-analysis of the neural correlates related to psychotherapy in depression and anxiety also reported changes in the frontal lobes, cingulate gyrus, and precuneus following treatments, and proposed that these neural differences result from changes that these therapeutic interventions create for self-related information processing (Messina, Sambin, Palmieri, & Viviani, 2013). Since we show that sustained weight-recovery includes neural differences related to social self-evaluation, treatment of AN may benefit by targeting this specific neurocognitive process. We also found that frontal regions were related to the EDE-Q global score during social self-evaluation for patient-participants, linking eating disorder symptomatology to this social process. In concert, these data suggest that both decreases in eating disorder symptoms and sustained weight-recovery from AN alter frontal brain activations during social self-perception.

Eating disorder clinical symptoms correlated with neural activations in response to self-stimuli for several of the ROIs that differed in the groups. Most commonly, clinical imaging studies include one clinical group and one comparison group, and symptoms in the clinical group are correlated with neural activations. When we combined the patient-groups, we found that neural activations in the dorsal anterior cingulate, right inferior frontal gyrus, and bilateral fusiform gyri correlated with the eating symptoms, confirming the relevance of these regions in the disease process. However, within each group, symptom scales were not correlated with neural activations, underscoring the importance of considering both the course of illness as well as current symptoms during assessments of AN. In this study, considering the course of illness led to more robust neurocognitive differences related to self-perception than measures of acute eating disorder symptoms.
One limitation of Social Identity-V2 task is that valence and behavioral response were confounded, both because participants (across all groups) agreed with few negative stimuli and disagreed with few positive stimuli, and because a much larger set of stimuli would be necessary to separate condition, behavior, and valence (See Supplemental Analysis 1). Interestingly, the insula and inferior frontal gyri, areas more activated in both patient groups compared to the controls during direct self-evaluations, were recently associated with negative self-evaluations in healthy subjects (Cabanis et al., 2013). Ambwani et al. (Ambwani et al., 2015) recently reported perceptual impairments related to positive stimuli in AN. Previously, we observed reduced neural activations in the precuneus and superior temporal sulcus following positive social interactions in both ill and weight-recovered women with AN (McAdams, Lohrenz, & Montague, 2015). Further exploration of whether differences in the behavioral responses and neurocircuitry engaged for both positive and negative self-related stimuli remains an important question for future studies.

Another limitation involves the variability of the clinical characteristics of the patient groups. Groups were defined primarily on body mass index during the last six months (AN-C) and two years (AN-WR). These criteria were selected because maintenance of weight within a healthy range reduces the mortality and morbidity of AN, with most relapses reported during the first two years after treatment (Strober, Freeman, & Morrell, 1997). Second, this criterion is more objective than individual patient reports of symptoms. Additionally, because this study focuses on evaluating neural differences related to cognitions about self-perception, examining individuals that are acutely malnourished (i.e. patients with AN immediately upon entry to a treatment program or severely underweight) would be inappropriate, as numerous cognitive problems, such as difficulties with set-shifting and central coherence, are observed following
starvation even in non-clinical samples (Pender, Gilbert, & Serpell, 2014). The goal of this study was to assess whether neurocognitive differences in self-perception contribute to the pathology of AN, a process best served by focusing on patients after nutritional stabilization, rather than examining the progressive loss of cognition associated with starvation. Nevertheless because on the day of the scan, six AN-C subjects had a normal BMI and six AN-WR subjects reported significant eating disorder symptoms, we excluded those individuals for a post-hoc analysis, and repeated the whole-brain ANOVAs, with similar findings (Supplemental Analysis 2). Similarly, although both clinical groups showed moderate and similar amounts of depression and anxiety, neural differences in these self-perception tasks were not related to either depression or anxiety, only the eating disorder symptoms. These results suggest neural changes in social self-evaluation are specifically related to eating disorder symptoms, and may be important in both the development of the disease and its clinical course.

In conclusion, neural regions were connected both with cognitive processes underlying self-perception as well as illness and recovery in AN, leading to new possibilities for neural and cognitive treatment targets. The differences in medial prefrontal cortex during self-relevance suggests a biological risk trait, a problem that might be approached with deep brain stimulation or transcranial magnetic stimulation, techniques just beginning to be explored in AN (Lipsman et al., 2013; McClelland, Kekic, Campbell, & Schmidt, 2015). Most excitingly, neural differences related to engagement of the salience network during social self-evaluation were found in women with sustained weight-recovery compared to both other groups, suggesting that successful long-term weight recovery from AN involves not only reductions in acute symptoms of an eating disorder but also changes in the neurocircuitry engaged for self-stimuli. More
research is needed to determine whether interventions that explicitly target social self-evaluation can improve outcomes from AN.
References


10.1016/j.comppsych.2011.09.003


Treasure, J., & Schmidt, U. (2013). The cognitive-interpersonal maintenance model of anorexia nervosa revisited: a summary of the evidence for cognitive, socio-emotional and
interpersonal predisposing and perpetuating factors. *J Eat Disord, 1*, 13.


FIGURE LEGENDS

Figure 1. The top row (A) shows the MPFC-dACC, centered on its peak MNI coordinates, from the main effect of group ANOVA for the Self–Friend contrast of the Social Identity-V2 task, with the parameter estimates (β values) extracted from this ROI, averaged for each group, on the right. The bottom row (B) shows the MPFC-Cing, centered on its peak MNI coordinates, from the main effect of group ANOVA for the Self-Agree – Self-Disagree contrast of the Social Identity-V2 task with its parameter estimates, averaged for each group and plotted on the right.

Figure 2. Three clusters were identified in the main effect of group ANOVA for the Reflected – Self contrast of the Social Identity-V2 Task. The upper row (A) is centered on the peak response in the RIFG, with the dACC cluster also visible in the middle panel and the LIFG cluster, as it extends into the insula, also visible in the far right panel. In the bottom row (B), parameter estimates for the Reflected – Self contrast were extracted for each ROI, and the average for each group is shown.

Figure 3. Two clusters were identified in comparison of the AN-C – AN-WR groups in the F-Self – F-Other contrast of the Faces task. In the top row (A), a cluster in the left fusiform (LFusi) extending into the temporal lobe showed more activation in the AN-C group for F-Self images than in both the AN-WR and HC groups. In the bottom row (B), a cluster in the right fusiform (RFusi) extending into the temporal lobe also showed more activation in the AN-C group for F-Self images than in both the AN-WR and HC groups.
**TABLES**

**Table 1. Participant Characteristics**

<table>
<thead>
<tr>
<th></th>
<th>HC (N = 19)</th>
<th>AN-C (N = 22)</th>
<th>AN-WR (N = 18)</th>
<th>$F_{2,56}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>27.9</td>
<td>6.0</td>
<td>27.6</td>
<td>7.6</td>
<td>29.6</td>
</tr>
<tr>
<td>ED Onset Age (years)</td>
<td>-</td>
<td>17.4</td>
<td>6.0</td>
<td>14.0</td>
<td>3.7</td>
</tr>
<tr>
<td>Intelligence Quotient (WASI)</td>
<td>122.53</td>
<td>9.0</td>
<td>117.8</td>
<td>8.3</td>
<td>118.5</td>
</tr>
<tr>
<td>Current Body Mass Index</td>
<td>22.5</td>
<td>2.4</td>
<td>17.6</td>
<td>1.5</td>
<td>22.8</td>
</tr>
<tr>
<td>Eating Attitudes Test (EAT)</td>
<td>3.3</td>
<td>3.9</td>
<td>39.0</td>
<td>17.9</td>
<td>15.7</td>
</tr>
<tr>
<td>EAT-Dieting</td>
<td>1.7</td>
<td>2.9</td>
<td>20.4</td>
<td>10.9</td>
<td>10.1</td>
</tr>
<tr>
<td>EAT-Bulimia</td>
<td>0.4</td>
<td>0.8</td>
<td>9.5</td>
<td>4.4</td>
<td>3.2</td>
</tr>
<tr>
<td>EAT-Oral</td>
<td>1.3</td>
<td>1.6</td>
<td>8.1</td>
<td>5.9</td>
<td>2.0</td>
</tr>
<tr>
<td>Quick Inventory of Depression</td>
<td>1.6</td>
<td>2.1</td>
<td>6.7</td>
<td>5.3</td>
<td>5.1</td>
</tr>
<tr>
<td>Hamilton Anxiety</td>
<td>2.3</td>
<td>2.8</td>
<td>10.3</td>
<td>8.7</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Self-perception in anorexia nervosa
### Eating Disorder Examination Questionnaire (EDE-Q)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>3.94</th>
<th>1.44</th>
<th>2.14</th>
<th>1.36</th>
<th>4.03&lt;sup&gt;d&lt;/sup&gt;</th>
<th>&lt;0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDE-Q - Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDE-Q - Restraint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDE-Q - Eating Concern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDE-Q - Shape Concern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDE-Q - Weight Concern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HC, healthy comparison women; AN-C, women with anorexia nervosa; AN-WR, women in weight recovery after anorexia nervosa.

ED, eating disorder; EAT, Eating Attitude Test; EDE-Q, Eating Disorder Examination Questionnaire.

Post-hoc statistical comparisons were performed using Bonferroni correction:

- <sup>a</sup> AN-C differ significantly from HC (<0.05).
- <sup>b</sup> AN-WR differ significantly from HC (<0.05).
- <sup>c</sup> AN-C differ significantly from AN-WR (<0.05).
- <sup>d</sup> t-value for the comparison of AN-C and AN-WR (df=38, <0.05).

Self-perception in anorexia nervosa
Table 2. Characteristics of all regions of interest from the whole-brain comparisons and regressions.a

<table>
<thead>
<tr>
<th>Condition and Region</th>
<th>Neural ROI Characteristics</th>
<th>β Values from ROI</th>
<th>Group Comparisons using β-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume (mm$^3$)</td>
<td>Cluster Size</td>
<td>Peak Z</td>
</tr>
<tr>
<td>Medial Prefrontal Cortex (MPFC-dACC)</td>
<td>10718</td>
<td>232</td>
<td>4.34</td>
</tr>
<tr>
<td>Medial Prefrontal Cortex (MPFC-Cing)</td>
<td>6283</td>
<td>136</td>
<td>4.01</td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus / Insula (LIFG)</td>
<td>2356</td>
<td>51</td>
<td>3.95</td>
</tr>
<tr>
<td>R Inferior Frontal Gyrus/ Insula (RIFG)</td>
<td>3650</td>
<td>79</td>
<td>4.29</td>
</tr>
<tr>
<td>Dorsal Anterior Cingulate (dACC)</td>
<td>3603</td>
<td>78</td>
<td>4.07</td>
</tr>
<tr>
<td>L Fusiform / Temporal (LFusi)</td>
<td>5960</td>
<td>129</td>
<td>3.78</td>
</tr>
<tr>
<td>R Fusiform/ Temporal (RFusi)</td>
<td>4712</td>
<td>102</td>
<td>3.57</td>
</tr>
</tbody>
</table>

Self-perception in anorexia nervosa
### EDE-Q Global Regression for Patient-Participants Only using Social Identity-V2: Reflected – Self Contrast

<table>
<thead>
<tr>
<th>Region</th>
<th>Voxel Count</th>
<th>Z value</th>
<th>Peak MNI Coordinates</th>
<th>Z value</th>
<th>Peak MNI Coordinates</th>
<th>Z value</th>
<th>Peak MNI Coordinates</th>
<th>Z score</th>
<th>P value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorsal Anterior Cingulate (dACC-EDEQ)</td>
<td>2495</td>
<td>4.52</td>
<td>-3 4 32 40</td>
<td>-1.49</td>
<td>-1.49</td>
<td>0.06</td>
<td>-1.49</td>
<td>8.8</td>
<td>&lt;0.001</td>
<td>c,d</td>
</tr>
<tr>
<td>Left Middle Frontal Gyrus (LMFG)</td>
<td>3280</td>
<td>4.13</td>
<td>-36 28 36</td>
<td>-0.83</td>
<td>-0.83</td>
<td>0.68</td>
<td>-0.83</td>
<td>9.0</td>
<td>&lt;0.001</td>
<td>c,d</td>
</tr>
<tr>
<td>Right Middle Frontal Gyrus (RMFG)</td>
<td>1894</td>
<td>4.36</td>
<td>48 20 40</td>
<td>-0.90</td>
<td>-0.90</td>
<td>0.13</td>
<td>-0.90</td>
<td>6.6</td>
<td>0.003</td>
<td>c,d</td>
</tr>
</tbody>
</table>

All Clusters at $P_{FWE} < 0.05$. The volume of each voxel is 46.2 mm$^3$. Peak Z Score is at the MNI coordinates for the specified anatomical location. HC, healthy comparison women, AN-C, currently-ill women with anorexia, and AN-WR, weight-recovered women with anorexia. Effect sizes computed from Cohen’s $d$ using means of extracted $\beta$ values.

- AN-C differ significantly from HC ($P < 0.05$).
- AN-WR differ significantly from HC ($P < 0.05$).
- AN-C differ significantly from AN-WR ($P < 0.05$).
- df for the Faces Task was 2, 52.

Self-perception in anorexia nervosa
Figure 1. Differences Related to Self-perception and Self-relevance in Anorexia Nervosa

A

B

Figure 1. The top row (A) shows the MPFC-dACC, centered on its peak MNI coordinates, from the main effect of group ANOVA for the Self – Friend contrast of the Social Identity-V2 task, with the parameter estimates (β values) extracted from this ROI, averaged for each group, on the right. The bottom row (B) shows the MPFC-Cing, centered on its peak MNI coordinates, from the main effect of group ANOVA for the Self-Agree – Self-Disagree contrast of the Social Identity-V2 task with its parameter estimates, averaged for each group and plotted on the right.

165x123mm (300 x 300 DPI)
Figure 2. Three clusters were identified in the main effect of group ANOVA for the Reflected – Self contrast of the Social Identity-V2 Task. The upper row (A) is centered on the peak response in the RIFG, with the dACC cluster also visible in the middle panel and the LIFG cluster, as it extends into the insula, also visible in the far right panel. In the bottom row (B), parameter estimates for the Reflected – Self contrast were extracted for each ROI, and the average for each group is shown.

165x123mm (300 x 300 DPI)
Two clusters were identified in comparison of the AN-C – AN-WR groups in the F-Self – F-Other contrast of the Faces task. In the top row (A), a cluster in the left fusiform (LFusi) extending into the temporal lobe showed more activation in the AN-C group for F-Self images than in both the AN-WR and HC groups. In the bottom row (B), a cluster in the right fusiform (RFusi) extending into the temporal lobe also showed more activation in the AN-C group for F-Self images than in both the AN-WR and HC groups.