Theory of Mind and the Brain in Anorexia Nervosa: Relation to Treatment Outcome

Martin Schulte-Rüther, Ph.D., Verena Mainz, Ph.D., Gereon R. Fink, M.D., Beate Herpertz-Dahlmann, M.D., Kerstin Konrad, Ph.D.

Objective: Converging evidence suggests deficits in theory-of-mind (ToM) processing in patients with anorexia nervosa (AN). The present study aimed at elucidating the neural mechanisms underlying ToM-deficits in AN. **Method:** A total of 19 adolescent patients with AN and 21 age-matched controls were investigated using functional magnetic resonance imaging during performance of a ToM-task at two time points (in-patients: admission to hospital and discharge after weight recovery). Clinical outcomes in patients were determined 1 year after admission. **Results:** Irrespective of the time point, AN patients showed reduced activation in middle and anterior temporal cortex and in the medial prefrontal cortex. Hypoactivation in the medial prefrontal cortex at admission to hospital (T1) was correlated with clinical outcome at follow-up. **Conclusions:** Hypoactivation in the brain network supporting theory of mind may be associated with a social–cognitive endophenotype reflecting impairments of social functioning in anorexia nervosa which is predictive for a poor outcome at 1-year follow-up. J. Am. Acad. Child Adolesc. Psychiatry, 2012;51(8): 832–841. **Key Words:** anorexia nervosa, superior temporal cortex, social cognition, medial prefrontal cortex, theory of mind

Anorexia nervosa (AN) is characterized by a markedly low body weight, intense fear of gaining weight, and body image distortion. Although not central to the diagnostic criteria of AN, emerging evidence suggests additionally deficits in key aspects of social functioning. Patients appear to be socially withdrawn, and they report having smaller social networks, less social interactions,[1] and a reduced number of close friends.[2] There is also evidence for premorbid social problems, such as increased levels of loneliness, feelings of inferiority, and shyness,[3] and comorbidity with anxiety disorders, such as social phobia.[4] Furthermore, a certain amount of overlap between AN and autism spectrum disorders (ASD) has been suggested.[5] Both conditions share a common profile of rigidity, aloofness, and social disengagement,[6] and show similar patterns of neurocognitive dysfunction including impaired set-shifting,[7] weak central coherence,[8] and impaired theory-of-mind (ToM) abilities.[9] Conversely, a lower mean body mass index as well as disturbed eating behavior has been described in some ASD patients.[10] ASD is often accompanied by impaired ToM abilities, which can be defined as the metacognitive capability of understanding mental states of other people, such as beliefs, wishes, and desires. First behavioral studies have investigated ToM abilities in patients with AN[9,11] and found deficits particularly in acute patients. These patients suffer from profound starvation associated with hormonal dysregulation, a general decrease of performance in cognitive tasks, and reductions in gray matter volume.[12] It remains to be elucidated whether impairments in ToM functioning are fully reversible,[11] or persist after longer periods of recovery.[13] Persisting functional deficits in ToM might be detected with more sensitivity using brain-imaging methods even when behavioral studies[11] fail to reveal such effects. For example, studies in patients with attention-deficit/hyperactivity disorder (ADHD) consistently show brain hypoactivation also in circumstances in which behavioral measures do not differ between patients and controls.[14]

Supplemental material cited in this article is available online.

Emerging evidence relates a negative long-term outcome of AN to a history of poor social functioning (e.g., empathy or social interaction problems) at or before the onset of the disorder.[15–17] Reduced brain activity in ToM networks might be a sensitive predictor for clinical outcome in patients with AN. To investigate this issue further, we used functional magnetic resonance imaging (fMRI) in patients with AN at admission to hospital (T1) and discharge from hospital following weight recovery (T2). Clinical outcome was assessed at a 1-year follow-up. We used a modified version of the social attribution task (SAT)[18] that has been adapted for optimal sensitivity in fMRI investigations[19] to reveal differences in the neural mechanisms underlying ToM relative to healthy controls and to correlate clinical outcome with brain activation patterns. We expected reduced activation in brain networks underlying ToM, including medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), superior temporal sulcus (STS), and temporal pole (TP). These brain regions have consistently been implicated in ToM processing[20] and have been shown to be hypoactivated in patients with ASD.[21] Furthermore, we hypothesized that reduced activation in these brain areas might be predictive for a worse clinical outcome.

METHOD

Participants

Nineteen female adolescents (12–18 years old) diagnosed with AN according to *DSM-IV* criteria were recruited from the inpatient service of the Department of Child and Adolescent Psychiatry, University Hospital Aachen. All patients underwent a multimodal treatment program including nutritional rehabilitation and weight management, cognitive-behavioral therapy on an individual and group basis, and family-based interventions.[22,23] Symptoms associated with the eating disorder were assessed with a structured clinical interview for the assessment of anorexia nervosa and bulimia nervosa (SIAB-EX), the Eating Disorder Inventory (EDI, a self-report questionnaire), and Morgan-Russell scales.[24] Depression symptoms were assessed using the Beck Depression Inventory (BDI). The Toronto Alexithymia Scale (TAS-26) was used as a measure of alexithymia, and the Interpersonal Reactivity index (IRI) was used as a measure of self-rated empathy. Thirteen patients experienced the restrictive subtype of AN, whereas six patients met the criteria for the binge/purging subtype. One patient was medicated with diazepam and olanzapine (at T1), one patient was medicated with fluoxetine (at T2), and one patient was medicated with olanzapine and fluvoxamin (at T2). At follow-up, none of the patients received medication. One patient fulfilled the diagnostic criteria for obsessive-compulsive disorder (predominantly obsessive thoughts, F42.0) and was not excluded, because obsessive thoughts and obsessive personality traits are characteristic of AN.[25] Two patients were diagnosed with a moderate depressive episode (F32.1), which was considered to be related to the eating disorder. No patient or control participant had a history of substance abuse.

Twenty-one healthy female control participants (group-matched for age, overall IQ, and educational level) without a history of any psychiatric disorder were recruited via a local advertisement. All participants gave written informed assent, and their parents or caregivers gave written informed consent after a complete and detailed description of the study. The study was approved by the local ethics committee of the University Hospital Aachen, in accordance with the Declaration of Helsinki. Using an overlapping sample of patients, clinical data and analyses related to structural brain changes have been published elsewhere.[26]

Time Course of Measurements

For patients, T1 occurred 17.3 (SD = 8.9) days after admission to hospital, and T2 took place at discharge from hospital (107.1 [SD = 39.8] days after admission). At discharge from hospital, patients had reached a mean target weight corresponding to the 17th age-specific body mass index (BMI) percentile (SD = 8.2; 3rd percentile at T1 [SD = 4.7]). Healthy control participants were also investigated at two time points (average time from T1 to T2: 213.4 (SD = 138.3) days). Clinical outcome in patients was determined 1 year after admission to hospital ($n = 16$, dropout ($n = 3$) because of noncompliance).

Overall outcome was assessed using Morgan-Russell outcome scales which are generally used when judging outcome in AN and have documented reasonable to good psychometric properties.[24] The Morgan-Russell Average Outcome Score (MRAOS) is derived from a guided interview assessing core clinical features of AN including food intake, menstrual state, mental state, psychosexual adjustment and vocational adjustment. Scores are rated on a continuous scale by an experienced clinician, reduced to five subscales, and further averaged to provide a general estimate of the clinical status or outcome.

Experimental Paradigm

Participants viewed 15.1-second videos of three white geometric figures (triangle, circle, and diamond) moving against a black background.[19] Twenty-four videos were presented that belonged to either of three conditions (eight videos per condition) (Videos S1, S2, and

S3, available online, provides sample videos). "ToM" scenes were designed to suggest contingent interactions of the figures that can be interpreted easily as social. The participants were instructed to decide at the end of the video whether all shapes were "friends" or not. Two types of non-ToM videos were used: figures that were circling at various speeds around the box, reminiscent of little "bumper cars"; and figures that performed physical movements, i.e., following simple trajectories. The task for non-ToM videos was to decide whether all shapes were equally "strong," based on the trajectories and speed after collisions. Each video was preceded by a 3-second condition-specific instruction cue. Between videos, a fixation cross appeared for 12 seconds. Yes-or-no responses were given via a button press at the end of each video (Supplement 1, available online, provides further details).

MR Technical Parameters

MRI measurements were taken using a 3-Tesla TRIO magnetic resonance (MR) scanner (Siemens, Erlangen, Germany) with a standard circularly polarized (CP) head coil. For functional imaging, gradient-echo echoplanar T2*-weighted images (EPI) were acquired (time to echo [TE] = 30 milliseconds, repetition time [TR] = 2424ms, α = 90°, field of view [FOV] = 200 mm, slice thickness = 3.0 mm^3, matrix size = 64 × 64, 40 transversal slices, ascending slice acquisition) in one session (~12 minutes). Anatomical images were acquired using a T1-weighted 3D magnetization-prepared, rapid acquisition gradient echo (MP-RAGE) pulse sequence (TE = 3.03 milliseconds, TR = 2250 milliseconds, time for inversion [TI] = 900 milliseconds, α = 9°, FOV = 256 mm, voxel size = 1 × 1 × 1 mm^3, matrix size = 256 × 256, 176 sagittal slices, slice thickness = 1 mm).

Behavioral and fMRI Data Analysis Before and After Weight Rehabilitation

Data for the analysis of T1 and T2 was available for 21 healthy controls and 19 AN patients. fMRI-data for some participants was missing at either T1 or T2 because of dropout or issues with fMRI data quality (missing data at T1: nAN = 1, n_{HC} = 1; missing data at T2: nAN = 1, n_{HC} = 5; available data at T1: nAN = 18, n_{HC} = 20; available data at T2: nAN = 18, n_{HC} = 16). For behavioral data, the software package PASW Statistics 18 (SPSS Inc., Chicago, IL) was used. Data referring to the description of the subject sample (demographical variables, diagnostic parameters and questionnaire data) were analyzed with t tests. Behavioral data of the functional paradigm were analysed with the mixed linear models module (a flexible generalized linear model [GLM] framework for mixed designs allowing for missing data of repeated measures). Two-tailed inference was performed, unless otherwise indicated.

Functional data were analysed with SPM5 (Wellcome Department of Imaging Neuroscience, London, UK) implemented in MATLAB 7 (Mathworks, Natick, MA). After realignment, functional images were normalized into the Montreal Neurological Institute (MNI) coordinate space and smoothed with an 8 × 8 × 8 mm^3 Gaussian kernel (full-width half-maximum). Boxcar functions (aligned with the videos) were convolved with a canonical model of the hemodynamic response and its first-order temporal derivative. Movement parameters were included as additional regressors. Parameter estimates of the resulting general linear model were calculated for each voxel and each regressor. For population inference, a second-level random effects analysis was performed (analysis of variance [ANOVA]), using the factors condition (within-subjects, 3 × 2 regressors for each type of video and measurement), and group (between-subjects). Specific effects at each voxel were tested by applying the appropriate linear contrasts to the parameter estimates. Because an initial assessment of results related to the difference between both non-ToM conditions revealed no effects with respect to group or time, these conditions were collapsed. Therefore, all reported results related to the factor task pertain to the comparison between the ToM and the combined non-ToM conditions, resulting in a three-factorial ANOVA design (group [AN, control] × time [T1, T2] × task [ToM, non-ToM]).

For the main effect of tasks (across time points and/or groups), a strict voxelwise threshold was applied ($p < .05$ familywise error (FWE) correction, extent threshold $k > 30$). For the relevant task-by-group and task-by-group-by-time interactions the voxel level threshold was set at $p < .005$ ($t > 2.60$). To control for false positive results, a spatial extent threshold was used resulting in a cluster-level threshold of $p < .05$, corrected for multiple comparisons. For region-of-interest (ROI) analyses, small volume corrections were applied across each respective region ($p < .05$, voxel-level FWE-correction). Anatomical ROIs of superior/middle temporal gyrus and TP were constructed using the Wake Forest University (WFU) Pickatlas software. Because the location of brain activation related to ToM at the TPJ and MPFC is not clearly defined anatomically, functional ROIs for these areas were constructed (10-mm sphere) using the coordinates of TPJ given in a recent meta-analysis on empathy, ToM, and attention[27] and the coordinates of the mPFC given in a recent study that used a similar paradigm.[28] An additional analysis was performed that excluded medicated patients; however, this did not change the pattern of results. To exclude the possibility that morphometric changes in gray matter (GM) volumes induced by starvation might contribute to group differences in the fMRI analysis, we conducted an additional analysis using the biological parametric mapping toolbox (BPM)[29] implemented in

TABLE 1 Change in Diagnostic Variables From Admission to Hospital (T1) to Discharge From In-patient Treatment (T2), and Follow-up in Anorexia Nervosa Patients

	T1 n = 18		T2 n = 18		Follow-up n = 15	
Variable	Mean	SD	Mean	SD	Mean	SD
BMI	15.3	1,5	18.1	1.0	17.5	1.5
EDI	260.7	57.3	241.8	63.4	254.0	80.5
SIAB-EX	65.1	24.5	41.3	18.6	54.1	26.9
BDI	20.3	11.1	9.5	8.6	12.6	12.1
MRav	5.71	1.26	6.46[a]	1.40	7.41	2.12

Note: Follow-up took place 1 year after T1. BDI = Beck Depression Inventory; BMI = body mass index; EDI = Eating Disorder Inventory; MRav = Morgan Russell average score; SIAB-EX = structured interview for eating disorders (expert rater).
[a]Missing data: n = 1.

SPM5 for those contrasts that yielded significant group differences (Supplement 1, available online).

Clinical Outcome and Dysfunction in ToM Networks in AN

A regression analysis was performed to relate reduced brain activation during the acute phase (at T1) to clinical outcome at follow-up. The individual contrast images of the comparison ToM versus non-ToM at T1 was used as a measure of individual ToM-related brain activation and correlated with average Morgan-Russell scores at follow-up. Of the original sample, data from 15 patients was available for this analysis (n = 3 drop-out at T3, n = 1 no neuroimaging data at T1). To restrict the results to brain structures involved in theory of mind processing as measured by our functional task, a functional ToM-ROI was defined using the fMRI data from the control sample (at T1, contrast ToM vs. non-ToM) thresholded at $p < .005$, FWE corrected. Because the functional ROI comprised all our relevant areas of interest, no additional anatomical ROI analyses were performed. For this regression analysis we report results that survived a threshold of $p < .05$ (cluster level corrected for multiple comparisons (ROI), $p < .001$ voxel level). Again, an additional BPM regression analysis was performed to control for potentially confounding reductions in gray matter (Supplement 1, available online).

RESULTS

Demographical and Questionnaire Data

Control participants and patients were closely matched for age (AN: 15.7 [SD = 1.5]; controls: 15.8 [SD = 1.9]; $T_{38} = -0.166$, p = .869) and IQ (AN: 108.8 [SD = 8.5]; controls: 108.0 [SD = 16.1]; $T_{27.5} = 0.198$; $p = .844$). The BMI of control participants was within normal population range (T1: 22.7 [SD = 3.9]; T2: 22.8 [SD = 3.6]). Clinical data of the patients at T1, T2, and follow-up are summarized in Table 1, additional data on empathy and alexithymia scales are given in Table S1 and S2, available online. At the time of discharge (T2), a reduction in symptoms associated with the eating disorder could be observed for the patients (as compared to T1), as assessed with SIAB-EX ($T_{17} = 3.961$, $p < .001$), EDI ($T_{17} = 1.394$, $p < .091$) and Morgan Russell average score ($T_{16} = -1.440$, $p < .0846$). Furthermore, a decrease in depression symptoms, as assessed with the BDI ($T_{17} = 3.490$; $p < .002$) and a significant increase in BMI were observed ($T_{17} = -8.944$, $p < .001$, all p values one-tailed).

Behavioral Data of the Experimental Paradigm

The statistical analysis of percentage of correct responses revealed a main effect of condition (F (2,154.13) = 59.59, $p < .001$), which was due to a better performance during the ToM task as compared to both non-ToM tasks (pairwise post-hoc comparisons, corrected for multiple comparisons, $p < 0.001$) and a better performance for the physical task than the bumper-car task ($p < 0.05$) across all participants. Importantly, there were no effects or interactions between groups. With respect to reaction times, a significant main effect of condition could be observed ($F_{2,133.20} = 3.72$, $p < .05$) which was due to longer reaction times ($\Delta RT = 174$ milliseconds) for the ToM videos as compared to the physical videos ($p < .05$). There were no significant differences between the non-ToM conditions ($p > .99$, $\Delta RT = 40$ milliseconds) or between the ToM and the bumper-car condition ($p > 0.20$, δ RT = 134 milliseconds). Further-

TABLE 2 Peaks of Significant Group Differences

Anatomical Region	H	BA	x	y	Z	t	k
T1 and T2							
Middle temporal gyrus	L	21	−56	4	−20	6.56	457
Inferior temporal gyrus	L	20	−58	−14	−28	3.36	
Middle temporal pole	L	38	−52	14	−22	2.89	
Middle temporal pole	R	21	58	8	−20	5.31	405
Middle temporal gyrus	R	21	54	−6	−22	3.59	
Inferior temporal gyrus	R	20/21	56	−6	−30	3.11	
Superior temporal gyrus	R	22	66	−16	−4	4.60	412
Middle temporal gyrus	R	21	58	−38	−4	3.17	
Medial prefrontal cortex[a,b]	R	10	4	60	20	3.35	
T1							
Superior temporal gyrus	R	22	66	−16	0	5.18	438
Middle temporal gyrus/temporal pole[a,c,d]	L	21	−58	4	−18	5.20	
Middle temporal pole[a,c]	R	21	58	8	−20	4.82	
T2							
Middle temporal gyrus	L	21	−56	4	−20	5.83	537
Inferior temporal gyrus	L	20	−56	−12	−28	4.34	
Middle temporal gyrus	L	20	−50	−14	−16	4.02	
Middle temporal pole[a,d]	L	21	−52	6	−18	4.43	
Middle temporal pole[a,d]	R	21	56	10	−22	4.18	

Note: Threshold $p < .05$ *corrected for multiple comparisons at the cluster level (voxel level:* $p < .005$, $t > 2.60$*). BA = Brodmann area; H = hemisphere; k = cluster size; L = left; R = right.*

[a]*Small volume correction for multiple comparisons (familywise error correction,* $p < .05$*, voxel level).*

[b]*Region of interest (ROI) based on a 10-mm sphere around [12 64 22], coordinates for activation in the medial prefrontal cortex reported by Moriguchi* et al.[28]

[c]*ROI based on superior and middle temporal gyrus.*

[d]*ROI based on the temporal pole; x, y, and z refer to Montreal Neurological Institute (MNI) coordinates of local peaks of activation for the interaction contrast (theory of mind [ToM]> non-ToM)*$_{\text{Healthy Control (HC)}}$ *> (ToM > non-ToM)*$_{\text{Anorexia Nervosa (AN)–patients}}$ *across time points (admission to hospital [T1] and discharge from in-patient treatment [T2]) and separately for each time point (T1, T2).*

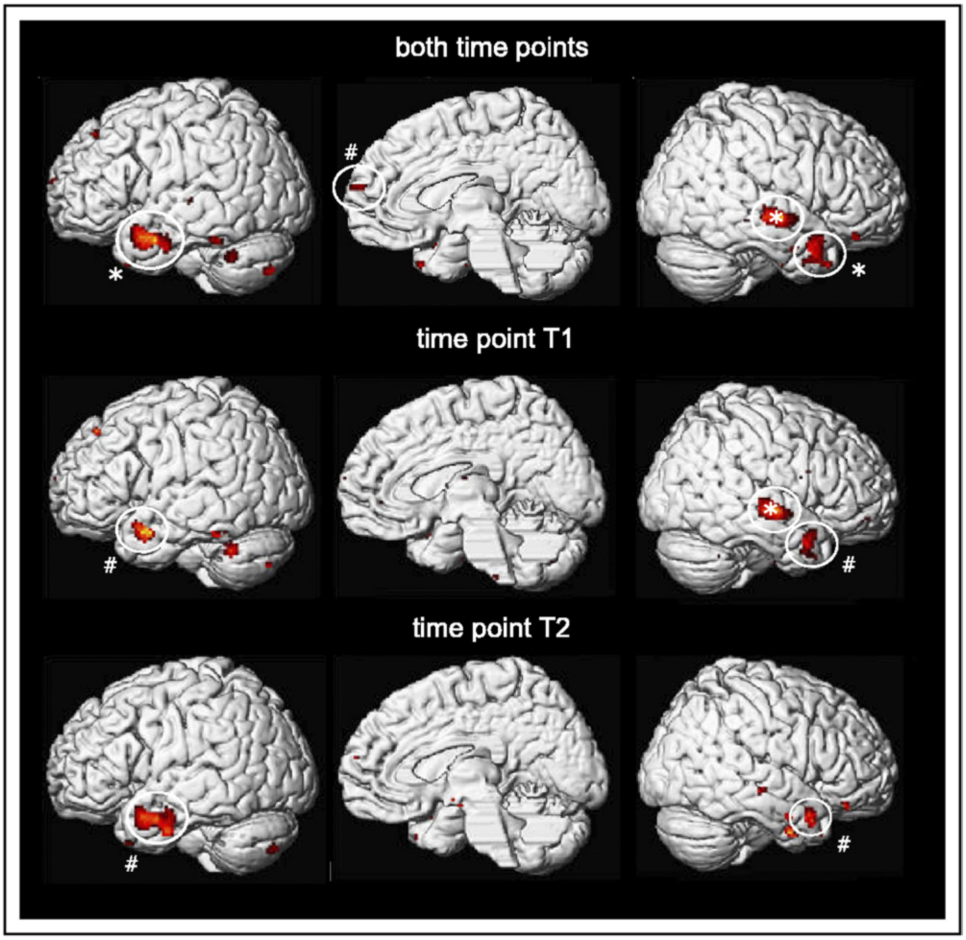
more, there was a significant effect of time ($F_{2,183.81} = 5.64$, $p < .05$) due to shorter reaction times (on average 142 milliseconds) at the second measurement (T2). Again, neither an effect of the factor group nor any interactions could be observed (Figure S1, available online).

fMRI Data

Across participants and time points, the comparison of ToM versus non-ToM tasks revealed neural activation in widespread brain areas including mPFC, superior and middle temporal gyrus, TP, fusiform gyrus, TPJ and precuneus (Figure S2 and Table S3, available online).

Interaction contrasts related to the effect of time (T1 versus T2) did not reveal any significant results (i.e., three-way interactions contrasts group-by-time-by-task and two-way interaction contrasts group-by-time). However, significant differences could be observed for a two-way interaction contrast (group-by-task), revealing reduced activation in patients with AN for the comparison between ToM and non ToM. This pattern could be observed in the right superior temporal gyrus, right TP, and left anterior middle temporal gyrus, extending into left TP and left inferior temporal gyrus (whole brain analysis; Table 2 and Figure 1). A ROI analysis in the mPFC revealed reduced neural activation in patients in the right mPFC. The reverse interaction contrast (testing for increased ToM-related neural activation in patients) did not reveal any significant differences. To further explore potential differences of both timepoints, we also assessed T1 and T2 separately for group-by-task interactions. In these analyses, the same pattern emerged, revealing significant interactions only for the HC > AN comparison. For T1, significant differences were observed in right superior temporal gyrus (whole brain), middle temporal gyrus and bilateral TP (ROI). For T2, differences emerged in the left middle temporal gyrus (whole brain) and right TP (ROI). In the mPFC, hypoactivation in AN could not be observed at a

FIGURE 1 Group differences in brain activation related to the theory of mind (ToM) task. Note: Statistical Parametric Maps depict the interaction contrast (ToM > non-ToM)$_{\text{healthy controls}}$ > (ToM > non-ToM)$_{\text{patients with anorexia}}$ thresholded at $p < .001$, uncorrected, for both time points, as well as across time points. Circled clusters were significant either at $p < .05$ (cluster-level corrected, whole brain)* or at $p < .05$ (familywise error voxel-level corrected for regions of interest [ROI])#. All peak activations were due to increased brain activation in healthy controls (in comparison to patients) for the contrast ToM versus non-ToM tasks. Figure S4, available online, illustrates β values in peak activated voxels.



corrected threshold for the separate comparisons of T1 and T2. However, at a more liberal threshold ($p < .005$, uncorrected) hypoactivation in the mPFC could be observed at both time points. The results from the additional BPM analysis (discussed in Methods) confirmed that these group differences were not due to morphometric differences in GM volumes (Supplement 1, Figure S2, available online).

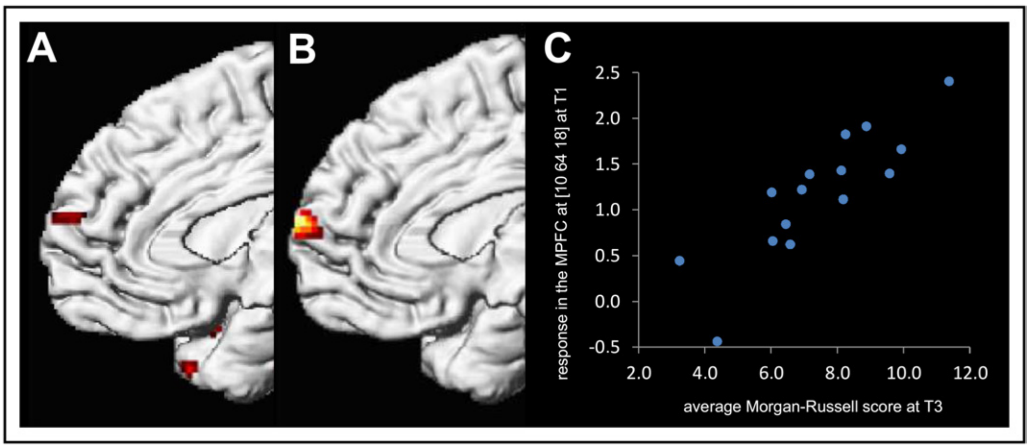
Clinical Outcome and Dysfunction in ToM Networks

Brain activation was positively correlated with Morgan-Russell scores in a cluster within the right mPFC (peak activated voxel: 10 64 18, k = 74, $p < .05$, cluster level corrected, similar to the peak of reduced activation in the group comparison; Figure 2), suggesting that reduced ToM-related brain activation at T1 is related to worse outcome. The results from the additional BPM regression analysis (discussed in Methods) confirmed that the observed correlation was not due to individual differences in GM volume reduction at T1 (Supplement 1, Results, available online).

DISCUSSION

The present study is one of the first reports on neural networks contributing to social cognition in patients with AN (see also McAdams and Krawczyk[30]) and the first to show a relation

FIGURE 2 Group differences and correlation with clinical outcome in the right medial prefrontal cortex. Note: (A) Statistical parametric map (SPM) depicts the peak voxels in the right medial prefrontal cortex (mPFC) that emerged in the interaction contrast (theory of mind [ToM] > non-ToM)$_{\text{healthy controls}}$ > (ToM > non-ToM)$_{\text{patients with anorexia,}}$ thresholded at $p < .001$ (voxel level), $p < .05$ corrected for multiple comparisons of regions of interest (ROI). (B) SPM displays the result of a regression analysis in anorexia nervosa (AN) patients demonstrating a correlation between less ToM-related brain activation (contrast ToM > non-ToM) at admission to hospital (T1) and worse outcome at follow-up 1 year later, thresholded at $p < .001$ (voxel level). The cluster in the right mPFC was significant at $p < .05$, corrected for multiple comparisons (ROI, refer to text for further details). (C) The scatterplot illustrates the correlation for the peak correlated voxel within the mPFC [10 64 18]. The y-values represent contrast estimates for the contrast ToM > non-ToM at T1; the x-values represent the individual average Morgan-Russell score of patients at follow-up. Significance of the correlation was determined in an unbiased voxel-wise regression-analysis, using correction for multiple comparisons.



between dysfunctional networks related to social cognition and treatment outcome.

Because the forced choice response of the fMRI-adapted social attribution task (SAT)[19] was not designed as a sensitive measure of individual or group performance differences, it is not surprising that we did not find behavioral differences between groups.[9,11] Similarly, reaction times did not differ between groups suggesting that differences in brain activation are unlikely to be related to starvation induced nonspecific performance deficits, but can be attributed to differences in the functional organization of brain networks supporting ToM abilities.

Across all participants, the ToM videos elicited activation in brain areas implicated in social cognition and ToM[19,20,28] such as mPFC, along the posterior and anterior STS, at the TPJ, TP, and in the precuneus, fusiform gyrus, amygdala, and inferior frontal gyrus (Figure S2, Table S3, available online). The medial prefrontal cortex, temporal cortex (including STS and TP) showed less activation in AN patients as compared to controls (Figure 1, Table 2). The STS is involved in the decoding of dynamic, socially relevant cues and has been implicated in diverse ToM paradigms.[20,31] Hypoactivation in patients with AN could be observed in the anterior aspect of the STS, which may reflect a deficit in the perception and identification of social cues as a precursor of ToM processing. Previous behavioral studies of deficits in ToM and emotional processing in AN[9,11] used paradigms that rely heavily on the correct decoding of subtle cues. Our results suggest that the observed deficits might arise because of functional impairments in brain areas that contribute to the extraction of socioemotional information. Less activation could also be observed in the anterior temporal cortex, including the TPs. The TPs have been suggested to provide a "semantic hub," binding together diverse aspects of semantic and autobiographical memory[32,33] and to integrate preprocessed perceptual inputs with emotional responses.[34] Such functions provide a frame of reference for mentalizing based on episodic and semantic recollections[35] and associated socioemotional responses. Patients with AN may tend to rely less on the representation of their aggregated social experiences when they infer social meaning from such abstract scenarios, suggesting impoverished memories or deficiencies in the contextual representations of socio-emotional situations. An alternative explanation could be an atypical perceptual processing style in AN, such as preoccupation with details.[8] Patients with AN may tend to focus on specific aspects of the videos rather than perceiving the clips as coherent social scenes. Interestingly, a detailed-focused cognitive style and weak central coherence have also been described in ASD.

AN was also associated with hypoactivation in rostral mPFC, which correlated with clinical outcomes after 1 year. This hypoactivation could be related to a mentalizing deficit associated with the disease (as suggested for ASD) or could be a consequence of aberrancies in the processing stream that provides input to an otherwise intact mentalizing module. It has been suggested that mPFC area constitutes the “core” region for representing mental states and anticipating the behavior of others.[20] In particular, anterior rostral mPFC is involved in mentalizing about oneself and individuals who are similar to oneself.[36] Reduced brain activation in this area has also been reported for conditions characterized by a deficit in emotional self-awareness (such as alexithymia),[28] which may also be associated with AN.[37] Increased alexithymia scores in AN were replicated in our sample. It might be speculated that patients with AN experience difficulties in relating abstract social scenarios to themselves. This conclusion is in accordance with the observed deficits in emotional ToM tasks in acute AN patients[9] and the persisting emotional processing deficits in recovered patients.[11]

Interestingly, prior work showed that GM loss in the closely adjacent anterior cingulate cortex (extending into anterior rostral mPFC) is correlated to an index of lifetime symptom severity in recovered women with AN.[38] For the present sample (which consisted of adolescents and primarily first-onset patients) GM alterations in mPFC did not contribute to the functional group effect. Whether early subtle functional aberrancies in mPFC, such as those observed in the present study, may become structural deficits after a longer period of illness needs to be investigated in future studies.

A recent study used a similar paradigm investigating adult participants in the process of recovery from AN,[30] including patients at different stages of recovery. This study reports similar hypoactivations in temporal brain areas (middle temporal gyrus, temporal pole), additional hypoactivation in TPJ, inferior frontal gyrus, and fusiform gyrus but no effects in medial prefrontal cortex. Whether these differences in comparison to our results are mainly due to age effects or differences in homogeneity of the patient group remains to be investigated in future studies.

Persisting Hypoactivation in Brain Networks Underlying ToM

Starvation has a profound impact on cognition and socio-emotional behavior, is associated with structural brain alterations[12] and hormonal dysregulation, and could therefore also result in changes of brain function. However, structural brain changes seem to be almost completely reversible after recovery and weight gain.[39] In our sample, we observed significant weight gain and recovery from AN symptoms, but no changes in brain activation patterns between admission (T1) and discharge (T2). Hypoactivation in brain areas related to ToM were still evident after weight recovery, irrespective of variations in GM density. Our data thus suggest that these persistent alterations in brain networks related to ToM may reflect a social-cognitive endophenotype associated with AN.[6] Recently, AN has been conceptualized as a neurodevelopmental disorder[40] with a neurocognitive profile similar to ASD[5] (i.e., impairments in set shifting, cognitive flexibility, central coherence, and theory of mind). The etiology of AN may be influenced by an interaction of genetic, hormonal, and psychosocial factors during childhood and adolescence that contribute to a maladaptive emotional, cognitive, and social development, facilitating the onset of disordered eating behavior.[40] Consistent with this notion, increased prevalence of depression and anxiety, including social phobia and separation anxiety,[4] as well as reduced social functioning can all be observed even before the overt onset of the eating disorder.[3,15] The concept of a social-cognitive endophenotype associated with AN is appealing; however, additional studies are needed to support this idea. Endocrine abnormalities were still evident at T2 in our sample ($n = 17$ of 19 patients still experienced primary or secondary amenorrhea) and might still have contributed to aberrant functional brain activation patterns. For example, estrogen levels (which are typically low in severely underweight patients) have been associated with changes in brain activation patterns for cognitive and emotional processing.[41] Future studies should therefore investigate the integrity of ToM networks after complete neuroendocrine recovery.

Hypoactivation in ToM networks and Prediction of Outcome

Good social functioning (including good family relationships) have been shown to be an important predictor of favouable outcome in AN.[15] Furthermore, premorbid poor social relating,[15] empathy deficits and social interaction problems[17] have been linked to poor outcome in AN. It remains unclear whether such findings reflect nonspecific psychosocial influences during the course of the disorder or whether they are indicative of a specific social-cognitive deficit involved in AN. The present study is the first to report a direct relationship between reduced activation of brain networks underlying ToM processing and poor clinical outcome, suggesting that dysfunction of the mPFC plays a key role for the course of the disorder. In accordance with our interpretation, it has been shown that impaired performance in abstract mentalizing tasks is correlated with poor global outcome in AN.[13] Furthermore, good "mentalization abilities" (i.e., the reflection about mental states) in social relationships may protect from the emergence of disordered eating behaviors.[42] These findings are in line with the idea of a social-cognitive endophenotype that represents a risk factor for the onset of AN and may result in a less favourable prognosis during the course of the disorder.

It should be noted that we have reported results based on a group-level analysis. fMRI group-level activation maps are highly reproducible in a repeated-measurements setting; however, future studies should assess whether social-cognitive paradigms provide sufficient test–retest reliability on the individual level[43] to allow the detection of subtle treatment-related effects. On the behavioral level, it might be helpful to use additionally open-ended scoring formats of verbal descriptions of the ToM videos to reveal behavioral effects related to the SAT.[18]

The identification of social-cognitive endophenotypes associated with poor outcome or the risk of developing AN may become a valuable strategy to develop specifically tailored interventions. Treatment of AN is often unsuccessful, and many patients relapse after initial weight recovery. In the light of emerging evidence for deficits in specific social-cognitive abilities such as ToM and emotion recognition,[9,11] social skills training or similar strategies to improve social cognition might prove an effective add-on to standard therapeutic strategies or may even allow for modulation of the disease at presymptomatic stages. It is conceivable that good mentalizing abilities play an important role during the process of recovery or may protect from relapse. Interestingly, a "mentalization-based psychotherapy"[44] and emotion skills training[45] have recently been proposed for the treatment of AN. Furthermore, it is of note that family-based therapy is among the most effective treatment approaches in adolescent patients with AN.[46] Family-based therapy focuses on parental influence to promote eating and weight control; it fosters communication between parents and child, and the enhancement of familial functioning as an essential element of successful therapy. Such socially driven treatments constitute an optimal framework for implicit social skills training, and one might speculate that this is one reason for their effectiveness in adolescents with AN. &

Accepted June 1, 2012.

Dr. Schulte-Rüther is with the University Hospital Aachen (UKA) and the Institute of Neuroscience and Medicine (INM-3) at the Research Centre Jülich. Dr. Mainz is with the Institute of Medical Psychology and Medical Sociology at UKA and the INM-3 at the Research Centre Jülich. Dr. Fink is with the INM-3 at the Research Centre Jülich and the University Hospital Cologne. Dr. Herpertz-Dahlmann is with the UKA and Jülich-Aachen Research Alliance (JARA) Translational Brain Research. Dr. Konrad is with the UKA, the INM-3 at the Research Centre Jülich, and JARA Translational Brain Research.

This study was supported by the Bundesministerium für Bildung und Forschung grant BMBF 01GV0602 (B.H.-D., K.K.).

We are grateful to all patients and their families for their participation. We thank Bob Schultz at the Center for Autism Research, Children's Hospital of Philadelphia, for providing the video stimuli of the social attribution task; Reinhild Schwarte and Melanie Krei at the Department of Child and Adolescent Psychiatry, Aachen, for their valuable help with study coordination and patient assessment; and our colleagues from the Department of Child and Adolescent Psychiatry at the University Hospital Aachen, from the Cognitive Neuroscience Section (INM-3) and Imaging Physics Section (INM-2) at the Research Centre Jülich.

Disclosure: Dr. Fink has served as an editorial board member of Cortex, Zeitschrift für Neuropsychologie and Fortschritte der Neurologie Psychiatrie. He has received royalties from the publication of the book "Funktionelle MRT in Psychiatrie und Neurologie" and "Neurologische Differentialdiagnose," has received honoraria for speaking engagements from Teva, GlaxoSmithKline, and Boehringer Ingelheim, and has received research support from the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft. Dr. Herpertz-Dahlmann has served as a consultant to Eli Lilly and Co. and has received industry research funding from Medice and Vifor. Dr. Konrad has served as an editorial board member of the Journal of Neural Transmission, the Journal of Child Psychology, and Zeitschrift für Kinder- und Jugendpsychiatrie, has received speaking fees from Eli Lilly and Co., Novartis, and Medice, has received industry research funding from Vifor, and has received research support from the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft. Drs. Schulte-Rüther and Mainz report no biomedical financial interests or potential conflicts of interest.

Correspondence to Martin Schulte-Rüther, Ph.D., Institute of Neuroscience and Medicine (INM-3, Cognitive Neuroscience), Research Center Jülich, 52425 Jülich, Germany; e-mail: mschulte@ukaachen.de

http://dx.doi.org/10.1016/j.jaac.2012.06.007

REFERENCES

1. Krug I, Penelo E, Fernandez-Aranda F, *et al.* Low social interactions in eating disorder patients in childhood and adulthood: a multi-centre European case control study [published online ahead of print April 5, 2012]. J Health Psychol. 2012;DOI:10.1177/1359105311435946.
2. Fairburn C, Cooper Z, Doll H, Welch S. Risk factors for anorexia nervosa - Three integrated case-control comparisons. Arch Gen Psychiatry. 1999;56:468-476.
3. Troop N, Bifulco A. Childhood social arena and cognitive sets in eating disorders. British J Clin Psychol. 2002;41:205-211.
4. Kaye W, Bulik CM, Thornton L, Barbarich N, Masters K. Comorbidity of anxiety disorders with anorexia and bulimia nervosa. Am J Psychiatry. 2004;161:2215-2221.
5. Gillberg C. The long-term outcome of childhood empathy disorders. European Child and Adolescent Psychiatry. 1996;56:52-56.
6. Zucker NL, Losh M, Bulik CM, *et al.* Anorexia nervosa and autism spectrum disorders: guided investigation of social cognitive endophenotypes. Psychol Bull. 2007;133:976-1006.
7. Robert M, Tchanturia K, Stahl D, Southgate L, Treasure J. A systematic review and meta-analysis of set-shifting ability in eating disorders. Psychol Med. 2007;37:1075-1084.
8. Lopez C, Tchanturia K, Stahl D, Treasure J. Central coherence in eating disorders: a systematic review. Psychol Med. 2008;38:1393-1404.
9. Russell TA, Schmidt U, Doherty L, Young V, Tchanturia K. Aspects of social cognition in anorexia nervosa: affective and cognitive theory of mind. Psychiatry Res. 2009;168:181-185.
10. Sobanski E, Marcus a., Hennighausen K, Hebebrand J, Schmidt MH. Further evidence for a low body weight in male children and adolescents with Asperger's disorder. Eur Child Adolesc Psychiatry. 1999;8:312-314.
11. Oldershaw A, Hambrook D, Tchanturia K, Treasure JL, Schmidt U. Emotional theory of mind and emotional awareness in recovered anorexia nervosa patients. Psychosom Med. 2010;72:73-79.
12. Kingston K, Szmukler G, Andrewes D, Tress B, Desmond P. Neuropsychological and structural brain changes in anorexia nervosa before and after refeeding. Psychol Med. 1996;26:15-28.
13. Gillberg IC, Billstedt E, Wentz E, *et al.* Attention, executive functions, and mentalizing in anorexia nervosa eighteen years after onset of eating disorder. J Clin Exp Neuropsychol. 2010;32:358-365.
14. Rubia K, Cubillo A, Woolley J, Brammer MJ, Smith A. Disorder-specific dysfunctions in patients with attention-deficit/hyperactivity disorder compared to patients with obsessive-compulsive disorder during interference inhibition and attention allocation. Hum Brain Mapp. 2011;32:601-611.
15. Strober M, Freeman R, Morrell W. The long-term course of severe anorexia nervosa in adolescents: survival analysis of recovery, relapse, and outcome predictors over 10-15 years in a prospective study. Int J Eating Disord. 1997;22:339-260.
16. Gillberg IC, Wentz E, Gillberg C, Anckarsäter H, Råstam M. Adolescent-onset anorexia nervosa: 18-year outcome. Br J Psychiatry. 2009;194:168-174.
17. Gillberg IC, Råstam M, Gillberg C. Anorexia nervosa outcome: six-year controlled longitudinal study of 51 cases including a population cohort. J Am Acad Child Adolesc Psychiatry. 1994;33:729-739.
18. Klin A. Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: the social attribution task. J Child Psychol Psychiatry Allied Disciplines. 2000;41:831-846.
19. Schultz RT, Grelotti DJ, Klin A, *et al.* The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Phil Trans R Soc London Series B Biol Sci. 2003;358:415-427.
20. Gallagher HL, Frith CD. Functional imaging of "theory of mind." Trends Cogn Sci. 2003;7:77-83.
21. Castelli F, Frith CD, Happé F, Frith U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain. 2002;125:1839-1849.
22. Herpertz-Dahlmann B, Salbach-Andrae H. Overview of treatment modalities in adolescent anorexia nervosa. Child Adolesc Psychiatr Clin N Am. 2009;18:131-145.
23. Herpertz-Dahlmann B, Hebebrand J. Assessment and Treatment of Eating Disorders and Obesity. In: Martin A, Scahill L, Kratchovil C, eds. Pediatric Psychopharmacology. Oxford: Oxford Univsrsity Press; 2011:570-586.
24. Morgan HG, Hayward AE. Clinical assessment of anorexia nervosa. The Morgan-Russell outcome assessment schedule. Br J Psychiatry. 1988;152:367-371.
25. Herpertz-Dahlmann B. Adolescent eating disorders: definitions, symptomatology, epidemiology and comorbidity. Child Adolesc Psychiatr Clin N Am. 2009;18:31-47.
26. Mainz V, Schulte-Rüther M, Fink GR, Herpertz-Dahlmann B, Konrad K. Structural brain abnormalities in adolescent anorexia nervosa before and after weight recovery and associated hormonal changes [published online ahead of print April 17, 2012]. Psychosom Med. DOI:10.1097/PSY.0b013e31824ef10e.
27. Decety J, Lamm C. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neurocientist. 2007;13:580-593.
28. Moriguchi Y, Ohnishi T, Lane RD, *et al.* Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. NeuroImage. 2006;32:1472-1482.
29. Casanova R, Srikanth R, Baer A, *et al.* Biological parametric mapping: a statistical toolbox for multimodality brain image analysis. NeuroImage. 2007;34:137-143.
30. McAdams CJ, Krawczyk DC. Impaired neural processing of social attribution in anorexia nervosa. Psychiatr Res. 2011;194:54-63.
31. Schulte-Rüther M, Fink GR, Markowitsch HJ, Piefke M. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. J Cogn Neurosci. 2007;19:1354-1372.
32. Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? The representation of semantic knowledge in the human brain. Nature Rev Neurosci. 2007;8:976-987.
33. Simmons WK, Martin A. The anterior temporal lobes and the functional architecture of semantic memory. J Int Neuropsychol Soc. 2009;15:645-649.
34. Olson IR, Plotzker A, Ezzyat Y. The enigmatic temporal pole: a review of findings on social and emotional processing. Brain. 2007;130:1718-1731.
35. Frith CD, Frith U. The neural basis of mentalizing. Neuron. 2006;50:531-534.
36. Mitchell JP, Macrae CN, Banaji MR. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron. 2006;50:655-663.
37. Bydlowski S, Corcos M, Jeammet P, *et al.* Emotion-processing deficits in eating disorders. Int J Eating Disord. 2005;37:321-329.
38. Mühlau M, Gaser C, Ilg R, *et al.* Gray matter decrease of the anterior cingulate cortex in anorexia nervosa. Am J Psychiatry. 2007;164:1850-1857.
39. Wagner A, Greer P, Bailer UF, *et al.* Normal brain tissue volumes after long-term recovery in anorexia and bulimia nervosa. Biol Psychiatry. 2006;59:291-293.
40. Connan F, Campbell IC, Katzman M, Lightman SL, Treasure JL. A neurodevelopmental model for anorexia nervosa. Physiol Behavior. 2003;79:13-24.
41. Dietrich T, Krings T, Neulen J, *et al.* Effects of blood estrogen level on cortical activation patterns during cognitive activation as measured by functional MRI. Neuroimage. 2001;13:425-432.
42. Rothschild-Yakar L, Levy-Shiff R, Fridman-Balaban R, Gur E, Stein D. Mentalization and relationships with parents as predictors of eating disordered behavior. J Nerv Ment Dis. 2010;198:501-507.
43. Plichta MM, Schwarz AJ, Grimm O, *et al.* Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. NeuroImage. 2012;60:1746-1758.
44. Skårderud F. Eating one's words: part III. Mentalisation-based psychotherapy for anorexia nervosa—an outline for a treatment and training manual. Eur Eating Disord Rev. 2007;15:323-339.
45. Davies H, Fox J, Naumann U, *et al.* Cognitive remediation and emotion skills training for anorexia nervosa: an observational study using neuropsychological outcomes. Eur Eating Disord Rev. 2012;20:211-217.
46. Lock J. Evaluation of family treatment models for eating disorders. Curr Opin Psychiatry. 2011;24:274-279.

SUPPLEMENT 1

METHOD

Participants and Diagnostic Assessments

Nineteen patients who had been referred to inpatient treatment for the first time with a mean duration of illness (time between first onset of symptoms as assessed with the structured Interview for Eating Disorders—expert rater [SIAB-EX] and admission to hospital) of 11.6 months (SD = 8.1 months) were included in the study and could be investigated at admission to hospital (T1) and discharge from in-patient treatment (T2). At follow-up, three patients decided not to participate.

A structured clinical interview was performed to assess co-morbid psychiatric diseases in patients and to exclude any psychiatric disease in healthy controls (Diagnostisches Interview Bei Psychischen Störungen Im Kindesalter [Kinder-DIPS][1]).

German versions of the Wechsler Intelligence Scale for Children (WISC-III/WISC-IV)[2] and the Culture Fair Intelligence Test (CFT-20R)[3] were used to assess overall IQ. In patients, IQ measurements were performed after an initial phase of weight gain to minimize the effects related to acute starvation.

In patients, symptoms associated with the eating disorder were assessed with SIAB-EX, a structured clinical interview for the assessment of anorexia nervosa and bulimia nervosa,[4] the eating disorder inventory (EDI,[5] a self-report questionnaire), and the Morgan-Russell criteria.[6] Depression symptoms were assessed using the Beck Depression Inventory (BDI).[7] Thirteen patients were characterized by the restrictive subtype of anorexia nervosa (AN), whereas 6 patients met the criteria for the binge/purging subtype. Assessment of overall outcome at follow-up was based on the Morgan-Russell scales.[8] We used the averaged scale score for regression analyses with brain activation. Patients completed the Toronto Alexithymia Scale (TAS-26)[9] as a measure of alexithymia and the Interpersonal Reactivity index (IRI)[10] as a measure of self-rated empathy at each time point, whereas controls completed these measures at either T1 or T2.

Healthy control participants did not take part in a follow-up investigation; however, they were also investigated at two time points. The average time from T1 to T2 for the healthy controls was 213.4 (SD = 138.3SD) days. Note: this interval is significantly different from the T1–T2 interval in patients. However, additional regression analysis in the control sample (correlating the individual time difference between T1 and T2 against brain activation at T2 (contrast theory of mind [ToM] > non-ToM) did not reveal any significant influence of time interval.

Treatment Program

All patients were treated using the same general treatment program, which included elements of different treatment approaches. The elements of the treatment program consisted of weight management and refeeding (regular weighing procedures, reinforcement based weight gain schedules, planned meals, control of physical exercise), training of eating behavior and nutrition (nutrition education, model based learning, educational cooking, successive rebuilding of self-managed eating, training of eating in social situations), psychotherapy (individual cognitive behavioral therapy, group therapy), and family based therapy (information about target weight and behaviors, support of familial communication, education of supporting behaviors, group based education on aspects of eating disorder, training of family eating).

Patient Comorbidities

One patient who fulfilled the diagnostic criteria for obsessive-compulsive disorder (predominantly obsessive thoughts, F42.0) was not excluded, because obsessive thoughts and obsessive personality traits are characteristic of AN.[11,12] Two patients were diagnosed with a moderate depressive episode (F32.1), which was considered to be related to the eating disorder.

Stimuli and Paradigm

To study the neural substrates of ToM processing, we used a modified version of the social attribution task[13] optimized for functional magnetic resonance imaging (fMRI) block design. Stimuli were identical to those used in a previous study[14] (supplementary videos SV1, SV2, and SV3 provide sample videos of each condition). The videos of the different conditions were designed to be equivalent with respect to general visual input, as well as movement speed, quantity, and location. A box was located in the center of the screen with one wall that could open and close like a door, allowing the shapes to enter. ToM videos were designed to suggest a sense of personal agency, reciprocal and contingent social interactions that can be interpreted easily as social. The participants were instructed to pay close attention to the interactions of the shapes and to decide at the end of the video whether all shapes were or were not "friends." In addition, two types of non-ToM videos were used: The first type of video depicted shapes that were circling at various speeds around the box. These were reminiscent of little "bumper cars," i.e., small racing cars, which had some playful collisions while circling. The other type of videos depicted shapes that performed physical movements. The shapes followed simple trajectories and were reflected after collisions with each other, the box and the border of the screen. The task for the two non-ToM types of videos was to pay close attention to the interactions and collisions of the shapes and to decide whether all shapes were equally "strong.". This decision should be based on the trajectories of the shapes and speed after each collision. All scenes were designed in such a way that the essential information providing the clue to the correct answer was given at the end of the film to ensure continuous attention to the film and to prevent early button presses during the clips. Furthermore, participants were carefully instructed to respond only at the end of the video, thus ensuring that participants were attending to the stimulus material. Button presses were performed with the index finger ("yes") and the middle finger ("no") of the right hand using an magnetic resonance (MR)–compatible response device. All participants practiced the task outside the scanner with videos of each condition (not shown in the main experiment) until they were familiar with the instruction cues and the experimental tasks. During the fMRI experiment, a thin-film transistor (TFT) display was used to project the stimuli to a set of mirrors mounted on the head coil, resulting in a visual field size of $\sim 5.2° \times 3.5°$. For stimulus presentation and response collection, the software Presentation 12 (Neurobehavioral Systems, Albany, CA; http://www.neurobs.com) was used.

With this paradigm it is impossible to distinguish the effects of the stimulus material from the effects of the instruction, i.e., to tease apart the perceptual and cognitive aspects of ToM processing. One possibility would be to include a condition with ToM stimuli that should only be judged on their physical aspects. However, pilot testing of the stimulus material (see also Schultz *et al.*)[14] indicated that this distinction is difficult to achieve. Participants reported that they still consciously perceived the ToM films as social stories if they were asked to judge physical aspects of the social scenes only. This is in line with other behavioral and neuroimaging studies reporting that (even in the absence of any explicit instruction) such conditions are likely to trigger ToM processing. Thus, we decided not to include such a condition in the present paradigm to have optimal power for the comparison of ToM related processing vs. non-ToM related processing.

Participant Sample and Data Analysis of Repeated Measures for Behavioral and Neuroimaging Data at T1/T2

The final sample, which entered into the behavioral and neuroimaging analysis of T1/T2 data, consisted of 21 healthy controls (HC) and 19 AN patients. However, because of drop-out and issues of fMRI data quality (e.g., artifacts or strong movement of participants) data for some participants was not available for one time point (T1: nAN = 1, $n_{HC} = 1$; T2: nAN = 1, $n_{HC} = 5$). We therefore used statistical methods for the statistical analysis of behavioral and neuroimaging data that allowed for the inclusion of participants with missing data at a certain time point (detailed below).

Behavioral Data Analysis

The software package PASW Statistics 18 (SPSS Inc., Chicago, IL) was used to analyse behavioral data, demographic variables, and questionnaire data of the T1/T2 dataset. Demographic variables and questionnaire data were analysed using one-sample and two-sample t tests. For behavioral data analyses (reaction times and percentage of correct choices), the mixed linear models module of PASW 18 was applied. This module implements a flexible generalized linear model (GLM) framework for mixed designs (random effects and repeated measures) and, in contrast to the standard repeated measures analysis of variance (ANOVA), allows for missing data in repeated measurements.[15] The percentage of "correct" choices and reaction times for correct choices were analysed using the factor subject as a random effect and the factors condition (ToM, non-ToM [bumper], non-ToM [physical]), time (T1, T2), and group (AN, HC) as fixed effects. To account for correlated repeated measures, covariance estimation was performed. All main effects and interactions were included into the models.

Standard fMRI Data Analysis

The first five volumes of each functional time-series were discarded to allow the MR signal to reach a steady state. The remaining 315 images were realigned (reference scan: 50^{th} image of each time course). After realignment (rigid body transformation), functional images were normalized into the Montreal Neurological Institute (MNI) coordinate space and re-sampled at $2 \times 2 \times 2$ mm^3. Normalization parameters were determined by applying the "unified segmentation" routine (as implemented in SPM5)[17] to each individual subject's mean echo planar imaging (EPI) image. Anatomical scans were normalized into MNI space using the same method. A high-pass cut-off filter of 240 seconds was used to account for low-frequency drifts in the imaging data. Box-car functions (corresponding to the onset and offset of each video) were convolved with a canonical model of the hemodynamic response and its first-order temporal derivative to compensate for timing differences in slice acquisition. To handle within-subject autocorrelations an approximate AR(1) model was estimated at omnibus F-significant voxels ($p < .001$, used globally over the whole brain). For population inference, the contrast estimates for the simple effect of each experimental condition were taken to the second level using the first regressor of the first-level hemodynamic response function (HRF) model as an estimate of response height, and a random effects analysis was performed (mixed ANOVA, factors: condition $\times$ group $\times$ subject). Such models in SPM5 allow for missing data in repeated measure designs. Departures from sphericity assumptions were accommodated using the nonsphericity correction in SPM5 (modeling of variance components). For this procedure, unequal variance was assumed for all factors, and nonindependence of data was assumed for the factor condition (repeated measures). Specific effects at each voxel were tested by applying the appropriate linear contrasts to the parameter estimates. Experimental conditions containing both non-ToM tasks were modeled separately. However, because the initial assessment of interactions between group and both non-ToM tasks did not reveal significant effects (neither at the whole brain level, nor for the ROIs), both non-ToM tasks were collapsed for subsequent data analysis focusing on group differences. Anatomical regions of interest (ROIs) of superior/middle temporal gyrus and temporal pole (TP) were constructed using the Wake Forest University (WFU) Pickatlas software.[17] Functional ROIs of the temporoparietal junction (TPJ) and medial prefrontal cortex (mPFC) were constructed (10-mm sphere) using the coordinates of TPJ given in a recent meta-analysis on empathy, ToM and attention[18] and the coordinates of the mPFC given in a recent study that used a similar paradigm to investigate individuals with alexithymia.[19]

Biological Parametric Mapping Toolbox

For the T1/T2 neuroimaging data additional analyses were performed for the contrast $(\text{ToM} > \text{non-ToM})_{HC} > (\text{ToM} > \text{non-ToM})_{AN}$ using the biological parametric mapping toolbox (BPM Version 1.5; Casanova *et al.* 2007) to exclude the possibility that morphometric differences between patients and controls might have influenced the observed group effects in fMRI data. Individual contrast images of the fMRI analysis (ToM versus non-ToM) were entered into a second-level model, that included the individual segmented gray matter images as voxelwise covariates. To obtain the individual anatomical covariates the toolbox VBM5.1 (http://dbm.neuro.uni-jena.de/vbm/download/), implemented in SPM5, was used. Individual T1 anatomical images were segmented into gray matter and white matter and spatially normalized into MNI space. The toolbox uses the unified segmentation approach[16] and standard processing parameters for implementation of voxel-based morphometry. The BPM-analyses were carried out for both time points separately (T1 and T2) and for the combined effect of both time points (T1 and T2). To optimize statistical power for this supplementary analysis, we focused only on those regions that were identified as showing stronger activations for controls (as compared to patients with anorexia) in the comparison between ToM and non-ToM tasks, which were the only significant group differences in the present study (Table 2 in main text). An ROI analysis was used to test for the specific influence of the anatomical covariate and the fMRI group difference within these regions.

For the regression analysis of neuroimaging data at T1 with diagnostic data at follow-up, an additional BPM regression analysis was performed. In addition to the regressor that was used in the standard analysis (average Morgan-Russell scores), an anatomical voxelwise covariate was added (see above).

RESULTS

Groups did not differ in empathic abilities as assessed by the IRI; however, patients scored higher on alexithymia scales (TAS-26) (Table S1). No changes could be observed in alexithymia scores or IRI scores, except for a significant increase at T2 for the subscale "personal distress" (Table S2).

Results of the additional analysis using the BPM toolbox indicated that group differences in the standard SPM analysis were due to true group differences in brain activation and could not be explained by the confounding influence of the anatomical covariate (see Figure S3). Significant differences only emerged with respect to the factor group ($p < .05$ false discovery rate [FDR]–corrected for ROI) in the same regions of temporal cortex as in the standard SPM-analysis (for all three analyses: T1, T2, and T1/T2 combined). There were no significant influences of the anatomical covariate on group differences in activation. Activation in the mPFC did not reach statistical significance when corrected for multiple comparisons (coordinate: [4 56 20] $p < .112$, FDR; $p < .003$ uncorrected; combined T1/T2 model). Influences of the anatomical covariate, however, were also not significant in the mPFC.

The results of the regression analysis using the BPM toolbox revealed virtually identical results as in the standard analysis, a significant cluster in the right mPFC (12 62 18, k = 62), for the positive correlation of ToM-related brain activation and average Morgan-Russell score.

The results of the comparison of both non-ToM tasks are depicted in Table S4. We observed significantly elevated brain activation in the inferior occipital cortex, cerebellum, and precuneus.

REFERENCES

1. Unnewehr S, Schneider S, Margraf J. Kinder-DIPS, Diagnostisches Interview bei psychischen Störungen Im Kindesalter. Göttingen: Hogrefe; 2008.
2. Petermann F, Petermann U. Hamburg Wechsler Intelligenztest für Kinder—IV (HAWIK-IV). Göttingen: Hogrefe; 2008.
3. Weiß R. Grundintelligenztest Skala 2—Revision. 4. ed. Göttingen: Hogrefe; 2006.
4. Fichter MM, Herpertz S, Quadflieg N, Herpertz-Dahlmann B. Structured Interview for Anorexic and Bulimic disorders for DSM-IV and ICD-10: updated (third) revision. Int J Eating Disord. 1998;24:227-249.
5. Thiel A, Jacobi C, Horstmann S, *et al.* [German version of the Eating Disorder Inventory EDI-2]. Psychotherapie Psychosomatik Medizinische Psychologie. 1997;47:365-376.
6. Morgan HG, Hayward AE. Clinical assessment of anorexia nervosa. The Morgan-Russell outcome assessment schedule. Br J Psychiatry. 1988;152:367-371.
7. Hautzinger M, Bailer M, Worall H, Keller F. Beck Depressions Inventar (BDI). Bern: Huber; 1994.
8. Ratnasuriya RH, Eisler I, Szmukler GI, Russell GF. Anorexia nervosa: outcome and prognostic factors after 20 years. Br J Psychiatry. 1991;158:495-502.
9. Kupfer J, Brosig B, Bröhler E. Toronto-Alexithymie-Skala-26 (TAS-26) Deutsche Version. Göttingen: Hogrefe; 2001.
10. Davis MH. A multidimensional approach to individual differences in empathy. JSAS Cat Select Doc Psychol. 1980;10:85.
11. Herpertz-Dahlmann B. Adolescent eating disorders: definitions, symptomatology, epidemiology and comorbidity. Child Adolesc Psychiatr Clinic North Am. 2009;18:31-47.
12. Jordan J, Joyce P, Carter F, *et al.* Specific and nonspecific comorbidity in anorexia nervosa. Int J Eating Disord. 2008;41:47-56.
13. Heider F, Simmel M. An experimental study of apparent behavior. Am J Psychol. 1944;57:243-259.
14. Schultz RT, Grelotti DJ, Klin A, *et al.* The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Phil Trans R Soc London Series B Biol Sci. 2003;358:415-427.
15. Wolfinger R, Tobias R, Sall J. Computing Gaussian likelihoods and their derivatives for general linear mixed. SIAM J Sci Comput. 1994;15:1294.
16. Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005;26:839-851.
17. Casanova R, Srikanth R, Baer A, *et al.* Biological parametric mapping: a statistical toolbox for multimodality brain image analysis. NeuroImage. 2007;34:137-143.
18. Decety J, Lamm C. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neurocientist. 2007;13:580-593.
19. Moriguchi Y, Ohnishi T, Lane RD, *et al.* Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. Neuroimage. 2006;32:1472-1482.

TABLE S1 Questionnaire Data for the Anorexia Nervosa and Control Groups

	HC Group		AN Group				
Variable	**Mean**	**SD**	**Mean**	**SD**	**t**	**df**	***p* (2-tailed)**
IRI subscales[a]							
Empathic concern	16.8	4.1	17.9	3.4	0.929	36	.359
Perspective taking	16.7	6.1	17.0	5.2	0.148	36	.883
Personal distress	9.7	5.3	11.4	3.2	1.186	36	.244
Fantasy scale	15.3	6.0	13.2	5.7	−1.118	36	.271
TAS-26[b]							
Overall score	2.11	0.40	2.61	0.53	3.272	35	.002
Identifying feelings	1.74	0.47	2.42	0.80	3.143	35	.003
Describing feelings	2.16	0.78	2.89	0.82	2.781	35	.009
Externally oriented thinking	2.51	0.55	2.61	0.61	0.508	35	.614

Note: Student's t tests for independent samples were applied to test for statistical differences between groups. AN = Anorexia nervosa; HC = healthy controls; IRI = Interpersonal Reactivity Index; TAS-26 = Toronto Alexithymia Scale.
[a]Missing data (n = 2).
[b]Missing data (n = 3).

TABLE S2 Change in Questionnaire Data From Admission to Discharge in Anorexia Nervosa (AN) Patients

	T1		T2				
Variable	**Mean**	**SD**	**Mean**	**SD**	**T**	**df**	***p* (2-tailed)**
IRI subscales[a]							
Empathic concern	17.9	3.4	18.0	2.8	.000	16	1.000
Perspective taking	17.0	5.2	18.3	3.3	−.494	16	.628
Personal distress	11.4	3.2	12.6	3.1	−2.324	16	.034
Fantasy scale	13.2	5.7	13.6	4.8	−0.336	16	.741
TAS-26[a]							
Overall score	2.61	0.53	2.54	0.65	0.802	16	.435
Identifying feelings	2.42	0.80	2.20	0.89	1.365	16	.191
Describing feelings	2.89	0.82	2.82	0.92	0.190	16	.851
Externally oriented thinking	2.61	0.61	2.69	0.65	−0.137	16	.892

Note: Student t tests for dependent data were applied to test for statistical differences between time points. IRI = Interpersonal Reactivity Index (20); TAS-26 = Toronto Alexithymia Scale (19).
[a]Missing data (n = 1) Means and SDs refer to all data available for that time point (n_{T1} = 18, n_{T2} = 18). Statistical inference refers to cases in which both time points were available (n = 17).

FIGURE S1 Behavioral data. Note: Percentage correct responses (A) and response times (B) for healthy control participants (green) and patients with anorexia nervosa (blue). Error bars represent the standard error of mean (SE). T1 = admission to hospital; T2 = discharge from in-patient treatment.

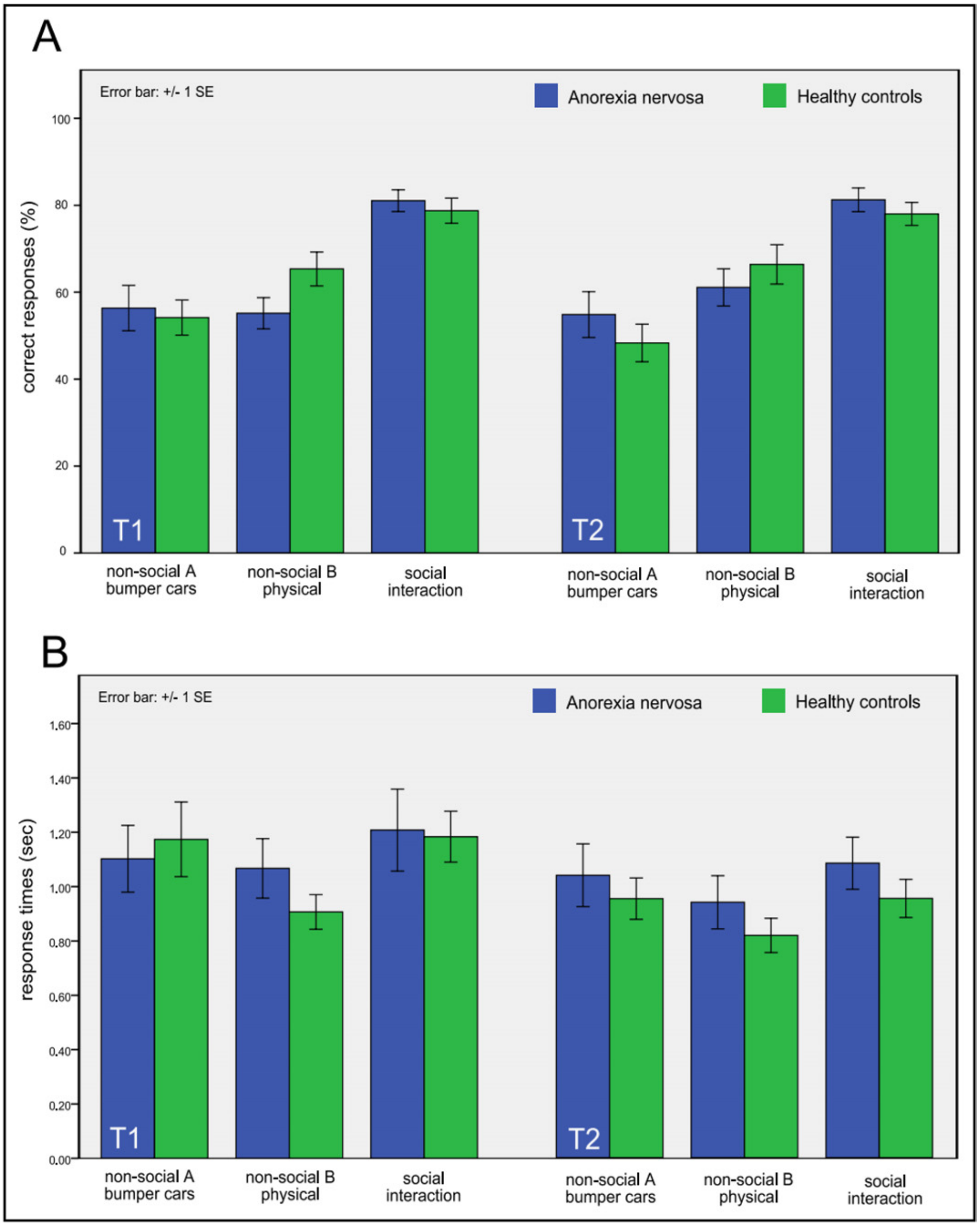


TABLE S3 Brain Activation for the Comparison of Theory of Mind (ToM) versus Non–Theory-of-Mind Tasks

			AN and HC				HC				AN			
Anatomical Region	**H**	**BA**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**
Middle temporal gyrus/STS	R	20	52	−16	−12	25.65	52	−16	−12	17.89	52	−16	−12	18.57
Middle temporal gyrus	L	20/21	−58	−54	8	14.95	−58	−52	−8	10.44	−58	−54	18	11.65
Middle temporal gyrus	R	21	54	0	−26	21.31	56	4	−24	15.74	54	0	−28	15.11
Middle temporal gyrus/TPJ	L	22	−48	−50	22	14.87	−48	−50	22	10.00	−50	−50	20	11.32
Middle temporal gyrus	R	21					56	−34	−4	14.10	48	−36	2	14.96
Middle temporal gyrus	L	37/39	−46	−58	16	15.19	−48	−60	16	10.04	−46	−58	18	11.93
Middle temporal gyrus	L	21/22	−54	−6	−18	14.76	−54	−4	−18	12.39	−58	−6	−14	9.38
Superior temporal gyrus/TPJ	R	22	64	−42	12	18.51	64	−42	12	12.68	60	−40	14	14.58
Superior temporal gyrus/TPJ	R	21	50	−56	22	17.79	50	−56	20	11.17	48	−56	22	14.47
Superior temporal gyrus	L	21/22	−62	−48	12	14.50	−58	−52	8	10.44	−58	−54	10	11.20
Supramarginal gyrus/TPJ	R	42	54	−46	24	18.65	54	−46	24	11.91	60	−48	24	15.48
Angular gyrus/superior temporalgyrus/TPJ	L	37/39	−46	−66	24	12.95	−48	−60	16	10.04	−40	−62	24	11.00
Angular gyrus	R	21/39	50	−56	22	17.79	50	−56	20	11.17	54	−58	24	14.09
Angular gyrus/middle temporal gyrus	L	21/22/39	−48	−50	22	14.87	−40	−54	18	12.39	−40	−62	24	11.00
Temporal pole	R	38	40	20	−32	18.00	42	18	−32	13.69	38	22	−32	12.28
Temporal pole	L	20	−44	10	−32	9.64	−42	12	−34	7.72	−44	10	−30	6.43
Temporal pole	L	20/38	−32	18	−34	11.33	−36	18	−36	8.63	−32	18	−32	8.17
Inferior temporal gyrus	L	20	−42	−14	−36		−42	−18	−34	6.96	−40	−16	−36	6.41
Fusiform gyrus	L	37	−46	−52	−20	15.39					−44	−52	−20	10.92
Fusiform gyrus/inferior temporal gyrus	L	20/37	−42	−42	−22	15.34	−46	−46	−20	11.87	−40	−44	−22	10.57
Fusiform gyrus/inferior temporal cortex	R	20/37	46	−42	−24	14.63	48	−44	−22	10.07	44	−42	−26	11.93
Fusiform gyrus	R	20	42	−14	−36	10.76	42	−14	−36	9.06	44	−16	−34	5.95
Precuneus	R		4	−56	36	15.56	4	−54	34	10.16	6	−56	36	12.49
Precuneus	L										−10	−52	38	8.47
Superior frontal gyrus pars medialis	R	10	6	54	16	15.75	6	54	18	12.33	10	60	12	12.01
Superior frontal gyrus pars medialis	R	9	8	52	36	14.56	8	52	38	10.21	8	50	36	11.02
Superior frontal gyrus pars medialis	L	9									−8	52	36	8.20
Orbitofrontal cortex		11	0	48	−26	11.60	0	50	−26	8.52	4	50	−22	8.71
Superior frontal gyrus	R	9	10	40	50	10.03	10	38	52	7.97	14	40	50	6.81
Supplementary motor area	R	6	10	16	64	12.80	8	18	62	9.03	10	16	64	9.56
Middle frontal gyrus	R	9	26	24	42	9.99	26	24	40	6.12	22	26	44	9.36
Precentral gyrus	L	6	−38	2	46	8.52					−40	4	48	6.37
Precentral gyrus/middle frontal gyrus	R	6/9	48	8	46	12.35	46	6	46	8.37	44	12	46	9.69
Middle frontal gyrus	R	9	26	24	42	4.91	26	24	40	6.12	22	26	44	9.36
Inferior frontal gyrus	R	45	52	32	−2	18.57	52	32	−2	12.65	52	32	0	14.12
Inferior frontal gyrus	R	45	56	30	12	17.71	56	30	12	12.12	54	32	10	13.49
Inferior frontal gyrus	L	44	−46	16	22	12.56	−50	18	20	8.37	−44	16	24	10.60
Inferior frontal gyrus	L	47	−50	30	−4	10.96	−46	32	−6	7.52	−52	30	−4	8.86
Cerebellum	R		16	−78	−44	12.02	16	−78	−44	9.22	18	−76	−44	7.89
Cerebellum	R		24	−76	−36	12.31	24	−78	−36	9.51	24	−76	−34	8.07
Cerebellum	L		−18	−78	−44	16.89	−20	−74	−34	11.34	−16	−78	−44	11.66
Cerebellum	L		−22	−78	−40	16.52	−24	−80	−38	13.49	−20	−74	−34	11.14

TABLE S3 Continued

			AN and HC				HC				AN			
Anatomical Region	**H**	**BA**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**
Amygdala	L		−24	−4	−26	9.47	−24	−4	−28	6.36	−24	−6	−26	7.26
Amygdala	L		−20	−6	−22	9.38	−18	−6	−22	6.91	−18	−8	−18	6.60
Amygdala	R		20	−6	−18	12.37	20	−6	−18	9.00	20	−6	−18	8.52
Thalamus	L		−6	−12	4	9.35	−6	−10	0	6.09	−6	−12	−4	9.13
Thalamus	R		6	−10	4	12.60	8	−10	4	9.08	6	−12	4	7.38

Note: Statistical Parametric Maps were thresholded at p < 0.001 corrected for multiple comparisons (familywise error corrected voxel level), extent threshold k >30; x, y, z refer to Montreal Neurological Institute (MNI) coordinates of local peaks of activation for the contrast ToM versus non-ToM, irrespective of time point. AN = anorexia nervosa; BA = Brodmann area; H = hemisphere; HC = healthy controls; L = left; R = right; STS = superior temporal sulcus; TPJ = temporoparietal junction.

FIGURE S2 Brain activation for the theory of mind (ToM) versus non-ToM tasks. Note: Statistical Parametric Maps (SPMs) depicting the contrast ToM > non-ToM for each group separately and for the combined contrast across participant group. Results are thresholded at $p < .05$ (familywise error [FWE] corrected, voxel level).

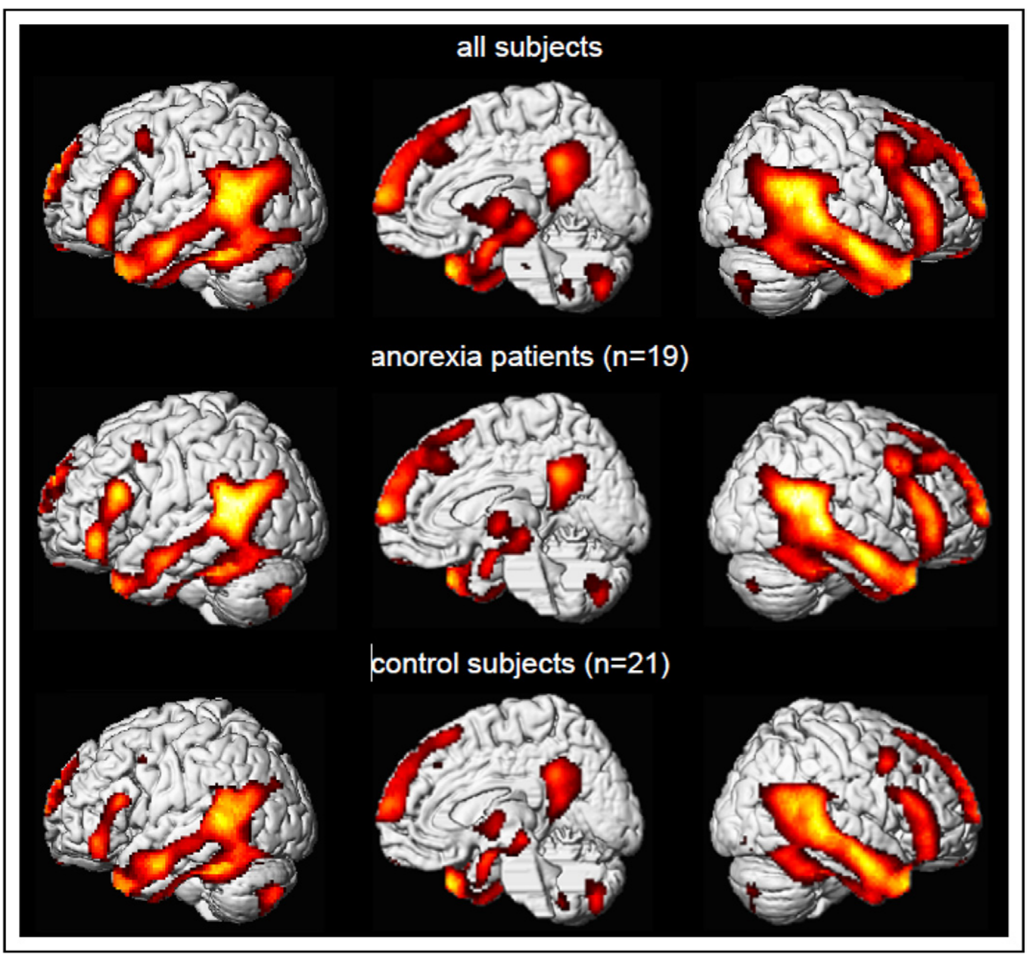


FIGURE S3 Effects of morphometric variations and group difference on the blood oxygen level dependent (BOLD) signal. Note: Statistical parametric maps (SPMs) depicting the results of an analysis of covariance (ANCOVA) performed using the biological parametric mapping toolbox[18] for both time points (Supplement 1, available online, provides details of the analysis). For illustrative purposes, SPMs are thresholded at $p < .05$ (uncorrected, voxel-level). Only the clusters for the group effect (irrespective of morphometric variations) were significant at $p < 0.05$ (voxel-level, corrected for multiple comparisons, region of interest [ROI]). T1 = admission to hospital; T2 = discharge from in-patient treatment.

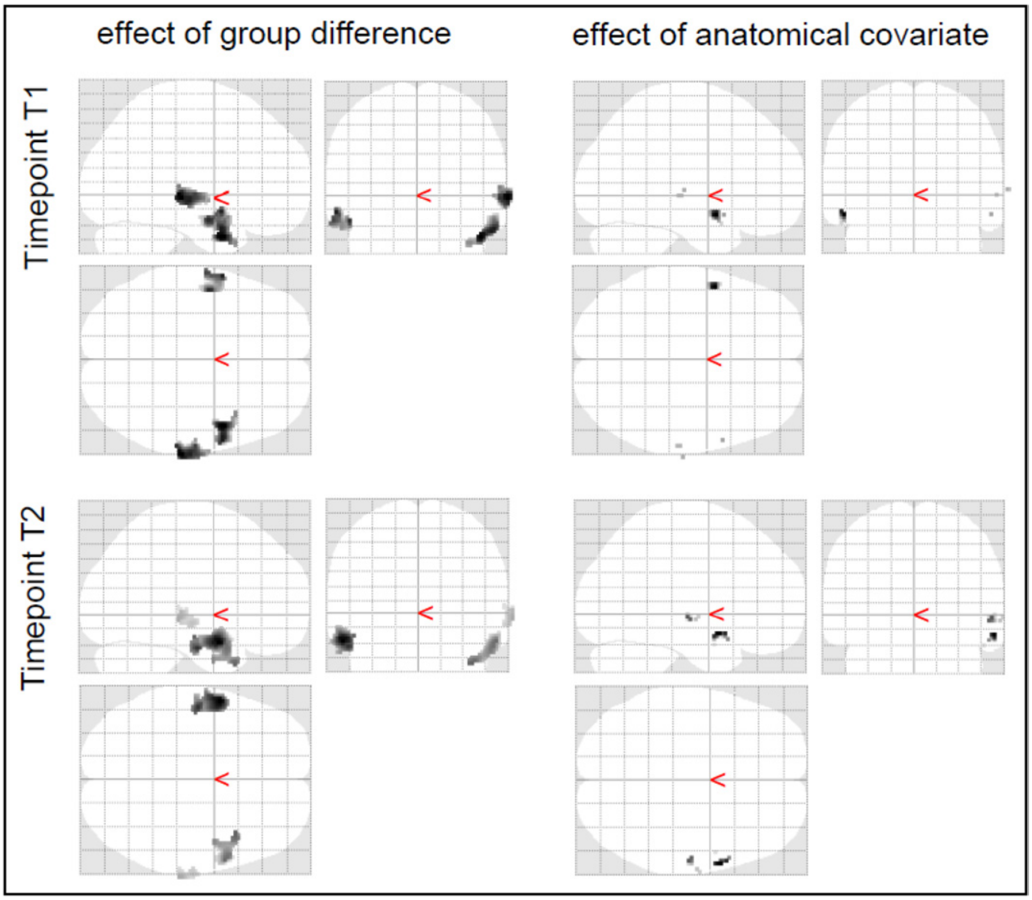


FIGURE S4 Contrast estimates of significant between group differences. Note: Mean contrast estimates for anorexia nervosa (AN) patients (in blue) and the control group (in green). Contrast estimates were calculated for peak activated voxels emerging in the interaction contrast (theory of mind [ToM] > non-ToM)$_{\text{healthy controls}}$ > (ToM > non-ToM)$_{\text{patients with anorexia}}$, across both time points (Figure 1, Table 3). Coordinates of are given in Montreal Neurological Institute (MNI) space. Error bars indicate 90% confidence intervals. Contrast estimates (contrast: ToM – non-ToM) were calculated separately for AN patients and healthy controls and both time points (admission to hospital [T1], discharge from in-patient treatment [T2]), respectively. Contrast estimates are in arbitrary units. They do not reflect an estimate of the effect size and are depicted here for visualization purposes. For statistical inference, a statistical parametric map (SPM) analysis was performed.

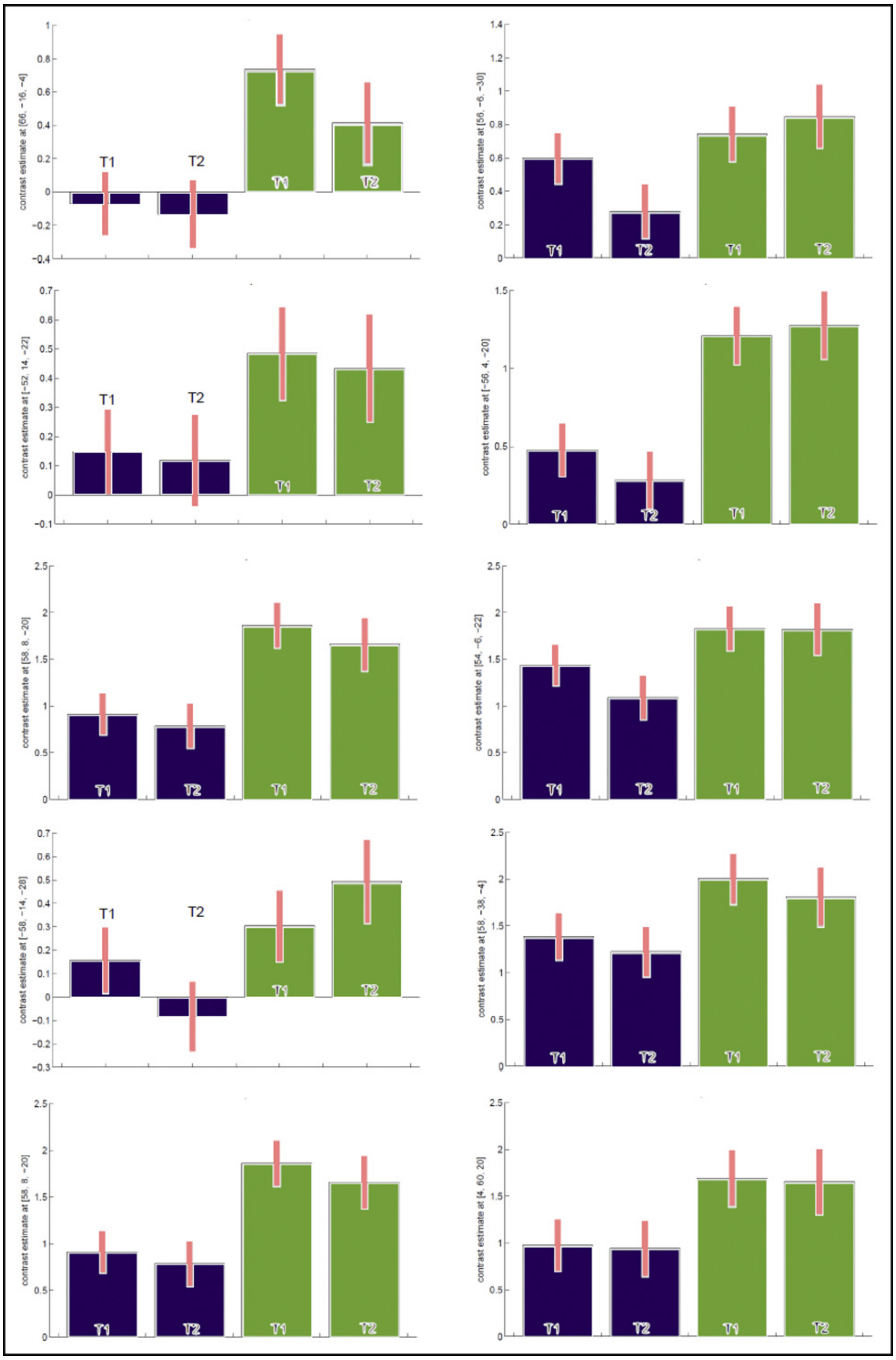


TABLE S4 Significant Activation for the Comparison of Bumper > Physical Task

			AN and HC				HC				AN			
Anatomical Region	**H**	**BA**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**	**x**	**y**	**z**	**t**
Bumper > physical														
Inferior occipital gyrus		18	−28	−98	−12	15.29	−28	−98	−12	11.07	−26	−100	−12	10.71
Inferior occipital gyrus		18	40	−92	−6	14.18	40	−92	−6	10.60				
Inferior occipital gyrus		17	26	−100	−8	14.10	26	−100	−8	9.81	26	−100	−8	10.14
Inferior temporal gyrus		37	52	−70	−12	7.81	56	−66	−12	6.16	48	−70	−12	6.78
Right postcentral gyrus		2	38	−30	46	6.75					38	−92	−6	9.85
Right postcentral gyrus		1	62	−10	38	6.58					62	−10	34	5.27
Precuneus			−20	−48	10	6.85								
Cerebellum							−28	−82	−38	5.77				

Note: Statistical Parametric Maps (SPMs) were thresholded at p *< 0.05 corrected for multiple comparisons at the voxel level (familywise error correction), extent threshold k >30; x, y, z refer to Montreal Neurological Institute (MNI) coordinates of local peaks of activation for the contrast non–theory of mind (ToM) [bumper] versus non-ToM [physical], irrespective of time point. AN = anorexia nervosa; BA = Brodmann area; H = hemisphere; HC = healthy controls; L = left; R = right.*