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 Patients appear to be socially withdrawn, and they report having smaller social networks, less social interactions,[1] and a reduced number of close friends.[2] There is also evidence for premorbid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