Neural Correlates of Theory of Mind Are Preserved in Young Women with Anorexia Nervosa

Leslie, Monica[a,b*];
Halls, Daniel[a*];
Leppanen, Jenni[a];
Sedgewick, Felicity [a,c];
Smith, Katherine[a];
Hayward, Hannah [d];
Lang, Katie [e];
Fonville, Leon [f];
Simic, Mima [g];
Mandy, William [h];
Nicholls, Dasha [i];
Murphy, Declan [d];
Williams, Steven [j];
Tchanturia, Kate [a, k]

Affiliations:

[a] King's College London (KCL), Institute of Psychiatry, Psychology and Neuroscience (IoPPN), Department of Psychological Medicine - London, United Kingdom
[b] Department of Psychology, University of Chester, Chester, United Kingdom
[c] School of Education, University of Bristol, Bristol, United Kingdom
[d] King's College London (KCL), Institute of Psychiatry, Psychology and Neuroscience (IoPPN), Department of Forensic and Neurodevelopmental Sciences - London, United Kingdom
[e] King's College London (KCL), Institute of Psychiatry, Psychology and Neuroscience (IoPPN), Department of Psychology - London, United Kingdom
[f] Division of Brain Sciences, Imperial College London
[g] South London and Maudsley NHS Foundation Trust
[h] Research Department of Clinical, Health and Educational Psychology, UCL, London, United Kingdom
[i] Division of Psychiatry, Imperial College London
[j] King's College London (KCL), Institute of Psychiatry, Psychology and Neuroscience (IoPPN), Department of Neuroimaging - London, United Kingdom
[k] Psychology Department, Illia State University, Tbilisi, Georgia

Corresponding Author: Professor Kate Tchanturia – kate.tchanturia@kcl.ac.uk - 103 Denmark Hill, Section of Eating Disorders, London SE5 8AF, United Kingdom

* Both authors contributed equally to the current manuscript.

Acknowledgements: KT would like to acknowledge: MRC-MRF Fund MR/S020381/1: BiomaRkers for AnorexIa NErvosa and autism spectrum Disorders- longitudinal study and MR/R004595/:1 The Triple A study (Adolescents with Anorexia and Autism: A search for biomarkers). DM is supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115300 (EUAIMS) and 777394 (AIMS-2-TRIALS), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme and Horizon2020 programs and the European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contribution.

The authors declare no conflict of interest.

Abstract

People with anorexia nervosa (AN) commonly exhibit social difficulties, which may be related to problems with understanding the perspectives of others, commonly known as Theory of Mind (ToM) processing. However, there is a dearth of literature investigating the neural basis of these differences in ToM and at what age they emerge. This study aimed to test for differences in the neural correlates of ToM processes in young women with AN, and young women weight-restored from AN, as compared to healthy control participants (HC). Based on previous findings in AN, we hypothesised that young women with current or prior AN, as compared to HCs, would exhibit a reduced neural response in the medial prefrontal cortex, the inferior frontal gyrus, and the temporo-parietal junction whilst completing a ToM task. We recruited 73 young women with AN, 45 weight-restored young women, and 70 young women without a history of AN to take part in the current study. Whilst undergoing a functional magnetic resonance imaging (fMRI) scan, participants completed the Frith-Happé task, which is a commonly-used measure of ToM with demonstrated reliability and validity in adult populations. In this task, participants viewed the movements of triangles, which depicted either action movements, simple interactions, or complex social interactions. Viewing trials with more complex social interactions in the Frith-Happé task was associated with increased brain activation in regions including the right temporo-parietal junction, the bilateral medial prefrontal cortex, the cerebellum, and the dorsolateral prefrontal cortex. There were no group differences in neural activation in response to the ToM contrast. Overall, these results suggest that the neural basis of spontaneous mentalising is preserved in most young women with AN.

Keywords: Anorexia nervosa; theory of mind; autism spectrum disorder; neuropsychology; functional magnetic resonance imaging

Anorexia nervosa (AN) is a severe eating disorder characterised by food restriction and compensatory behaviours leading to body weight which is excessively low for the individual's height and development status (American Psychiatric Association, 2013). AN has a complex aetiology, with a number of genetic and environmental risk factors contributing to onset of the disorder. Recent theoretical models have highlighted the importance of interpersonal difficulties in contributing to the onset and maintenance of AN (Schmidt & Treasure, 2006).

Theory of mind has (ToM) been defined as the ability to infer information about others' emotions, intentions, knowledge, and beliefs from social interactions or given information (Frith & Frith, 2005). ToM abilities are therefore critical in most social situations to effectively understand and respond to the behaviours and intentions of others. Problems in ToM have been well-documented in autism and recent research has also found problems in ToM among people with AN, including difficulties with emotional and cognitive ToM (Bora & Köse, 2016; Kerr-Gaffney, Harrison, & Tchanturia, 2019; Leppanen, Sedgewick, Treasure, & Tchanturia, 2018; Sedgewick et al., 2019). It is possible that ToM processes may hinder individuals' response to talking therapies, such as by contributing to poor self-insight, and may impact affected individuals' ability to access and utilise social support networks in the recovery process (Bora & Köse, 2016). It is, therefore, pertinent to better characterise the nature of ToM difficulties in AN and its underlying biological mechanisms in order to better understand the development of AN and possible social-cognitive targets for treatment intervention (Russell, Schmidt, Doherty, Young, & Tchanturia, 2009).

In the general population, ToM is associated with a complex network of brain regions. In particular, the temporo-parietal junction (TPJ) has been highlighted as a putative region that supports the formation of mental representations (Abu-Akel & Shamay-Tsoory, 2011; Döhnel et al., 2012). Following initial detection and representation of mental states, previous authors

have hypothesised that the TPJ subsequently relays this information via the superior temporal sulcus (STS) to limbic and paralimbic regions for emotional processing (Abu-Akel & Shamay-Tsoory, 2011; Gao et al., 2019).

The bulk of neuroimaging research administering a ToM task during scanning has found different patterns of activation in autistic people [1] compared to control participants. For example, research has found lower levels of activation in the right TPJ (Castelli, Frith, Happé, & Frith, 2002; Kirkovski, Enticott, Hughes, Rossell, & Fitzgerald, 2016) and altered functional connectivity between anterior and posterior brain regions among autistic people during ToM tasks (Kana, Keller, Cherkassky, Minshew, & Just, 2009). However, more recent evidence has suggested that these differences may be specific to men, with autistic women exhibiting similar activation in the right TPJ and ventromedial prefrontal cortex during a mentalising task compared to that of typically developing women (Kirkovski et al., 2016; Lai et al., 2019).

By contrast, recent evidence has highlighted differences in the brain networks recruited during ToM tasks in women with AN versus healthy controls, which may underpin functional differences in ToM abilities. McAdams and Krawczyk (2011), for example, found that, when compared to the healthy control participants, participants with a history of AN exhibited lower neural activation in brain regions forming part of the social cognition network, including the right inferior frontal gyrus, the bilateral TPJ, and the left fusiform gyrus during an implicit social attribution task. Schulte-Rüther, Mainz, Fink, Herpertz-Dahlmann, and Konrad (2012) conducted a later functional magnetic resonance imaging (fMRI) study using a similar ToM task in female adolescent inpatients with AN and healthy control participants. The authors found reduced neural activation in the middle and anterior temporal cortex and medial

[1] Identity-first language (i.e., autistic person), opposed to person-first language (i.e., person with autism), is preferred by many autistic people and their allies. Therefore, in this article, the authors use predominantly identity-first language to describe this population. Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442-462.

prefrontal cortex (mPFC) during the ToM task, as compared to the healthy control group. Furthermore, the level of hypoactivation in the mPFC was correlated with clinical outcome one year following discharge.

The current study aimed to expand on previous neuroimaging research into ToM in AN in a more highly-powered study, thus enabling us to draw more confident conclusions about the degree of difference in the neural underpinning of ToM in young women with AN, and those weight-restored from AN, as compared to age-matched controls. We were specifically interested in testing for differences in the neural correlates of ToM processes in young adults with AN and young adults in weight recovery from AN compared to healthy control participants. We also sought to investigate the relationship between ToM-related neural activation and autistic features in young women with AN. Whilst undergoing an fMRI scan, participants completed the Frith-Happé task, which is a commonly-used measure of ToM with demonstrated reliability and validity (Abell, Happe, & Frith, 2000; White, Coniston, Rogers, & Frith, 2011). Based on previous findings in AN, we hypothesised that the mPFC, the inferior frontal gyrus, and the TPJ would be associated with a reduced blood-oxygenated-level-dependent (BOLD) response in young women with, and weight-restored from, AN as compared to healthy controls, whilst completing the Frith-Happé task. We also hypothesised that greater levels of autistic characteristics in participants with AN would be associated with reduced ToM-related neural activation in the mPFC and related circuits extending to the TPJ.

Materials and Methods

Participants

A total of 188 young women between 16 and 25 years old participated in the current study. Seventy-three women met DSM-5 criteria for AN at the time of the study, 23 women were weight-restored from AN but exhibited continuing elevated levels of eating disorder symptoms, 22 women were in full recovery from AN, and 70 comparison women had no

current or prior history of an eating disorder. Given the low sample sizes for the weight-recovered and fully recovered participant samples, these groups' data were pooled into a single weight-restored participant group for all analyses. The BMI range of healthy control (HC) participants was 18.29 to 33.39, the BMI range of participants with acute AN (AAN) was 12.65 to 18.50, and the BMI range of participants weight-restored from AN (WR) was 18.36-26.81. The average duration of illness for participants with current AN was 3.10 years (*SD* = 2.56 years) and the average duration of illness for participants weight-restored from AN was 4.53 years (*SD* = 2.78 years). Demographic statistics and clinical characteristics associated with each participant sample are presented in **Table 1**. Full inclusion and exclusion criteria for the study and details of the participants' medication use are presented in the **Supplementary Material**.

Participants with AAN were recruited from the South London and Maudsley National Health Service Foundation Trust. The HC and WR participant groups were recruited via social media, via the website for BEAT (the UK's charity for eating disorders), and through advertisements in the local community. All participants provided written informed consent to take part in the study and, for participants under the age of 18, parental consent was also obtained. Ethical approval for the study was granted by the London – Surrey Borders Research Ethics Committee (REC Reference: 17/LO/0271). All study activities were in completed in accordance with the Declaration of Helsinki.

Measures

The Eating Disorders Examination – Questionnaire version (Fairburn et al., 2009), the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), the National Adult Reading Test (Nelson, 1982), Autism Quotient-10 item version (Allison, Auyeung, & Baron-Cohen,

2012), and the Autism Diagnostic Observation Schedule (Lord et al., 2000) were administered to the participants. Details about these measures are presented in the **Supplementary Material**.

The Frith-Happé Animations. The Frith-Happé animations depict a series of cartoons in which a red triangle and a blue triangle can be seen to move around a central open box, often in a way that implies they are animate and interacting (Abell et al., 2000). The Frith-Happé animations fall into three categories: 1) Random movement, in which the two triangles appear to float across the screen, occasionally bumping into each other, but displaying no symbolic social interaction; 2) Goal-directed movement, in which the triangles move in the same direction, and may appear to chase each other, but do not exhibit mentalising behaviour; and 3) ToM interactions, in which the triangles appear to take the other shape's thoughts and beliefs into account, such as by tricking or coaxing the other triangle to do something. The Frith-Happé animations are sensitive to difficulties with ToM even in autistic people who have an IQ within the normal range, who pass standard first- and second-order false belief tasks (Abell et al., 2000). The Frith-Happé animations have a standardised coding system that produces an accuracy measure and a language measure for each of the three types of trials. Each trial's accuracy is rated as 0 if the participant's narrative contains a plainly wrong description and/or focuses on an unimportant aspect, 1 if the participant's narrative contains a partial description of the sequence, but is imprecise or incomplete, and 2 if the participant's narrative is a spot-on description of the story or the actions represented. Each trial's language was coded as 0 if the participant describes a simple action with no interaction between the triangles, 1 if the participant describes interaction between the triangles without reference to mental states, and 2 if the participant uses mental state verbs to describe reciprocal interactions between the triangles.

Procedure

Each participant attended two study sessions. During the first session, participants completed the self-report questionnaires and structured clinical interviews (e.g., the ADOS). Participants were screened for MRI safety prior to proceeding to the second session.

Upon presentation to the second study session, participants completed a narrative version of the Frith-Happé animations outside of the scanner. During each trial of the Frith-Happé animations, the participants were asked to describe what they thought the triangles were doing. Participants' descriptions of each trial were audio-recorded and these behavioural data, collected outside of the scanner, were later analysed. Participants subsequently underwent an fMRI scan in which they completed a battery of neuropsychological tests. The Frith-Happé task was repeated inside the scanner as before except that, rather than the participant describing what the triangles were doing, at the end of each trial participants were instead asked to use a button box to indicate whether the triangles had exhibited random, goal-directed, or ToM movements. The multiple-choice version of the Frith-Happé animations has previously been validated in adults within the context of fMRI scanning paradigms (White et al., 2011).

fMRI Scan Acquisition

A total 307 volumes were acquired during the Frith-Happé task. Images within the fMRI scans were acquired with a slice thickness of 4mm and a slice gap of 0.5mm. A total of 28 slices were acquired in a top to bottom order. The field of view was 192mm^2 with a 64 x 64 matrix size. The resulting voxel size was therefore 3mm x 3mm x 4mm. The scan was conducted with an echo time of 30ms and a repetition time of 2,000ms. The flip angle was set to 80 degrees. A 3D high- spatial-resolution, Magnetisation Prepared Rapid Acquisition (3D MPRAGE) T1-weighted scan was also acquired. Field of view was 270mm^2, TR/TE/TI = 7.312/3.016/400ms. Two dummy scans were acquired at the start of the task and were subsequently discarded.

Statistical Analysis

Behavioural Data Analysis. The audio recording of each trial in the Frith-Happé task was coded by one researcher, and then checked by a separate researcher. Initial inter-rater reliability was 92.23%. Discrepancies were subsequently reviewed by the lead author, such that instances of agreement with the second coder were confirmed and instances of disagreement were resolved. We planned to compare the accuracy and language scores of the three participants groups for random, goal-directed, and ToM trials on the Frith-Happé task using between-groups ANOVAs in line with the analyses previously conducted by Abell and colleagues (Abell et al., 2000). However, as the residuals for the between-groups ANOVAs were not normally distributed, we instead conducted a Kruskall-Wallis test for each comparison and subsequently controlled for multiple corrections using an alpha rate of $p < .05_{\text{FWE-corrected}}$.

MRI Data Pre-processing. We conducted pre-processing of the MRI data using fMRIprep 1.2.6-1 (Esteban et al., 2017; Esteban et al., 2019), which is based on Nipype 1.1.7 (Gorgolewski, 2017; Gorgolewski et al., 2011). The full boilerplate associated with fMRIPrep, containing extensive details of pre-processing, is presented in the **Supplementary Material**.

MRI Data Analysis. We conducted both first- and second-level processing using FSL FEAT (FMRI Expert Analysis Tool) Version 6.00 (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004). At the single subject level, the data were modelled using the general linear model framework. We operationalised the "ToM" regressor as a linear contrast increasing in value from random trials, to goal-directed movements, to ToM trials and a separate contrast decreasing in social value from ToM trials, to goal-directed movements, to random trials. The BOLD signal was modelled by convolving our design matrix with a Double Gamma function. We included global signal, derivatives of motion parameters, squares of

motion parameters, and a scrubbing variable excluding volumes with a framewise displacement > 0.9 as confound variables at the single-subject level.

At the group level, we conducted region-of-interest (ROI) analyses using FSL featquery. We constructed the ROI masks based on peak coordinates from previous relevant studies. ROIs were 10mm spheres based on coordinates identified by previous ToM research for the right inferior frontal gyrus (MNI coordinates [52 28 8] (McAdams & Krawczyk, 2011)) and the right TPJ (MNI coordinates [54 -52 26] (Krall et al., 2015)). As a 10mm sphere localised in the mPFC crossed the brain boundary, we instead constructed a 9mm spherical mask within the mPFC in order to avoid extracting null data from outside of the brain (MNI coordinates [4 60 20] (Schulte-Rüther et al., 2012)). We subsequently conducted exploratory whole brain analyses using cluster level inference with a cluster threshold of $Z > 3.1$ and $p < .05$, corrected for multiple comparison using Gaussian random field theory.

Four participants did not complete the theory of mind task and two participants had scans of unusable quality, resulting in a total of 182 participants' data included in the final analysis. A power analysis conducted in G*Power revealed that our between-groups analyses were powered to detect small to medium effect sizes (f = 0.23) (Erdfelder, Faul, & Buchner, 1996).

Results

Behavioural Data Analyses

Descriptive statistics associated with the accuracy and language scores associated with the Random, Goal-Directed, and ToM trials for each participant group and results of the Kruskal-Wallis tests comparing the participant groups are presented in **Table 2**. There were no significant between-group differences in accuracy for any of the three trial types. The analysis

initially identified differences in the level of social language used for the random and theory of mind trials, such that HC participants tended to use greater levels of social language to describe random trials than participants with AAN and WR participants tended to use greater levels of social language to describe ToM trials than participants with AAN. However, these differences did not survive correction for multiple comparisons.

ROI Analyses

We conducted a between-groups ANOVA comparing mean BOLD activation within the mPFC, the TPJ and the inferior frontal gyrus. There were no significant differences between the three participant groups for any of the ROIs. We subsequently added psychiatric medication use as a covariate in a between-groups ANCOVA. This ANCOVA also did not reveal significant differences between the three participant groups for any of the ROIs.

Exploratory Whole-Brain Analyses

An initial one-sample *t*-test revealed 19 significant clusters associated with increasing complexity of the ToM contrast and a separate one-sample *t*-test revealed 20 significant activation clusters associated with decreasing complexity of the ToM contrast. These task-activated regions conform with previous norms reported within the theory of mind literature, including activation within the temporo-parietal junction, medial prefrontal cortex, and inferior frontal gyrus. The full results of these one-sample *t*-tests are presented in **Supplementary Table 1 and Supplementary Table 2.**

A between-groups ANOVA comparing the ToM contrast between the three participant groups did not reveal any significant clusters associated with increasing or decreasing social complexity of the ToM contrast. We next conducted a sensitivity analysis excluding participants taking psychoactive medication to account for any suppression of between-group

differences driven by psychotropic medication. This between-groups ANOVA also failed to detect any significant between-groups differences associated with increasing or decreasing complexity of the ToM contrast.

Finally, we conducted exploratory whole brain analyses within the AAN participant group including the AQ10, ADOS Communication subscale, ADOS interaction subscale, ADOS imagination and creativity subscale, the ToM accuracy and language scores, BMI, global EDE score, and illness duration as covariates in nine separate one-sample *t*-tests. The ADOS communication subscale and the ADOS interaction subscale were both correlated with BOLD response to decreasing complexity of the ToM contrast within the right extrastriate cortex (i.e., higher ADOS scores were associated with lower BOLD response to ToM trials). Cluster peaks for the ADOS communication subscale were located at MNI coordinates [23.5 -78.5 -10.5] and [15.5 -82.5 -16.5]. The cluster peak for the ADOS interaction subscale was located at MNI coordinate [23.5 -78.5 -12.5]. Illness duration was correlated with the BOLD response to increasing complexity of the ToM contrast in the left parahippocampal gyrus, MNI coordinate [-22.5 -22.5 -14.5] and to decreasing complexity of the ToM contrast in the left premotor cortex, MNI coordinates [-22.5 -4.5 55.5] and [-26.5 -0.5 63.5]. There were no significant associations between any of the other covariates and BOLD response to the ToM contrast amongst participants with current AN.

Discussion

The current study aimed to test for differences in the brain correlates of ToM processing in young women with AN, young women weight-restored from AN, and healthy comparison participants. We hypothesised that participants with, or weight-restored from, AN would exhibit reduced activation in the mPFC, the TPJ, and the inferior frontal gyrus in response to a ToM task, when compared to those without history of an eating disorder. However, the data

did not support any of these hypotheses, as there were no significant between-group differences in BOLD response to a spontaneous mentalising task. We also hypothesised that neural activation within the mPFC, the TPJ, and the inferior frontal gyrus would be negatively correlated with autistic traits amongst participants with AN. Our manipulation check revealed that task-activated regions conformed with previous norms reported within the theory of mind literature, including activation within the temporo-parietal junction, medial prefrontal cortex, and inferior frontal gyrus. The latter hypothesis was not supported by the results, as autistic traits were not associated with task-related activation in these three hypothesised regions. However, the ADOS communication and interaction scales were associated with task-related neural response in early visual processing regions. Furthermore, illness duration was found to be associated with task-related neural response in the left parahippocampal gyrus and left premotor cortex.

Our behavioural findings corresponded with previous studies which also found no evidence of differences in accuracy between women with a history of AN and healthy control participants on spontaneous mentalising tasks (McAdams & Krawczyk, 2011; Schulte-Rüther et al., 2012). However, the lack of group differences in brain response to the ToM task was an unexpected result, which contrasts with previous studies finding altered patterns of BOLD responses to a very similar task among adult women in recovery from AN (McAdams & Krawczyk, 2011) and in a previous study conducted in adolescents with AN (Schulte-Rüther et al., 2012). There are several potential explanations for this difference in findings. First, it may be the case that differences in the neural underpinning of ToM develop progressively throughout the course of the illness and remain for some time after recovery, which might explain why a different pattern of neural response to a similar ToM task has previously been observed amongst older adult women in weight recovery from AN, as compared to age-matched control participants, but not in our sample of young adults with AN (McAdams &

Krawczyk, 2011). However, this explanation does not account for the failure to replicate previously-documented differences in BOLD responding to a ToM task amongst adolescents with AN (Schulte-Rüther et al., 2012).

It is possible that the present results may reflect no true differences in the neural underpinnings of ToM across the entire population of young adults with AN. Indeed, our relatively large sample size of 188 young adults, including 73 young adults with current AN, 45 young adults in weight recovery from AN, and 70 healthy control participants, is likely to be associated with more stable effect sizes and reduced confidence intervals than the previous study conducted in young people with AN, which recruited only 19 participants with current AN. This is consistent with the notion that ToM impairments are present in a subgroup of those with AN, but do not feature on average across cases with adolescent onset (Stewart, McEwen, Konstantellou, Eisler, & Simic, 2017).

The current results suggest that differences are specifically observed in individuals with AN who are high in autistic characteristics. Specifically, higher levels of communication and interaction difficulties were associated with increased neural response to decreasing complexity of the ToM contrast. This finding may be explained by previous research demonstrating that, in contrast to neurotypical participants, autistic participants demonstrate a lack of attentional modulation when viewing social stimuli, which is associated with differences in the activation of early visual regions, including the primary visual cortex and extrastriate cortex (Bird, Catmur, Silani, Frith, & Frith, 2006). Previous evidence suggests that this between-groups effect is particularly pronounced for subtle, versus overt, social cues (Zürcher et al., 2013), which are exemplified by the representational social cues depicted by triangles in the Frith-Happé task.

The association between duration of AN and task-related activation in the left parahippocampal gyrus and left premotor cortex is, however, more difficult to explain on the basis of previous literature in populations with AN. In the general population parahippocampal gyrus activation has been observed in response to completing empathy and face recognition tasks (van Veluw & Chance, 2014; Völlm et al., 2006). It may be that the effects of more prolonged malnourishment disrupt circuits related to social memory and the perception of social stimuli mediated by the parahippocampal gyri. However, further evidence is needed to more clearly establish the functional significance of this finding.

The current findings add to our understanding of the complex pattern of differences exhibited by people with AN across different domains of ToM. A recent meta-analysis of ToM abilities in people with AN found that, while affected individuals exhibit statistically significant differences in the domains of emotional ToM, understanding simple social interactions, and understanding complex social interactions, there was no significant difference in the domain of implicit social attribution, measured in the current study (Leppanen et al., 2018).

Indeed, the extent of blanket differences in spontaneous mentalising abilities and gross differences in associated neural activation has more recently been questioned, even in autistic populations. For example, a recent large study recruiting more than 300 autistic participants found no differences in performance or neural activation on the Frith-Happé task, when compared to healthy control participants (Hayward et al., Unpublished results). Furthermore, previous evidence, which did find differences in the neural correlates of ToM in autistic men, did not find similar differences in autistic women (Kirkovski et al., 2016; Lai et al., 2019). These previous findings are difficult to reconcile with our current observation that some components of autistic traits are, indeed associated with neural response to a ToM task. Further research in large samples of autistic women will help to clarify whether such differences may

be specific to those with the greatest levels of communication and social interaction difficulties, as suggested by our current findings in young women with AN.

Strengths of the current study include the large sample size of women completing both behavioural and fMRI tasks, allowing greater confidence in the effect sizes found within the current set of analyses. However, this study is not without limitations, including the specific component of ToM measured within the Frith-Happé task. Thus, while the current study provided no evidence for differences in the brain underpinnings of spontaneous mentalising in young women with a history of AN versus healthy controls, problems in this population have previously been observed for other components of ToM (Leppanen et al., 2018), and may be associated with a different pattern of neural activation on other ToM tasks. Additionally, it is possible that presenting a descriptive version of the task prior to the neuroimaging scan resulted in a "training" effect, perhaps resulting in the recruitment of a greater degree of memory processes and lesser degree of theory of mind processes than would have been observed had participants viewed the task for the first time during the fMRI scan. Further research will therefore be required to corroborate these results and examine potential differences in the neural underpinnings of emotional ToM and complex social interactions. Finally, as this study was conducted exclusively in young women, the current findings should not be generalised to men or older adults with AN.

While the current study has replicated consistent findings in brain regions that underpin ToM processing, including within the rTPJ and mPFC, we did not find evidence for between-group differences in the neural underpinnings of spontaneous mentalising in young women with a history of AN versus healthy control participants. It should be noted that this null finding may be due, in part, to the specific ROI masks analysed in the current study. We based our ROIs on previous studies conducted in AN to maximise applicability to the population recruited in the current study. However, these previous activation peaks were observed in

relatively small samples, and a different pattern of results may have been observed had we based our ROIs on regions that are generally activated during the Frith-Happé task in HC participants. Future research will help to clarify whether different patterns of neural activation underpin behavioural performance in other domains of ToM and more clearly establish the functional significance of the association between illness duration and task-related neural response. Overall, the current set of findings suggests that the neural processing of spontaneous mentalising remains more intact in young women with AN than previously thought.

Table 1

Descriptive demographic and clinical statistics

	Healthy Control (*n* = 70)			Acute AN (*n* = 67)			Weight-restored AN (*n* = 49)			K-W Test Statistic	*p*
	Mean(SD)	Median(IQR)	Skew	Mean(SD)	Median(IQR)	Skew	Mean(SD)	Median(IQR)	Skew		
Age (Years)	19.64(3.30)	18.54(17.39-22.72)	0.52	18.70(2.78)	18.40(16.53-20.90)	0.44	19.72(3.27)	18.94(17.31-22.41)	0.23	3.85	.278
BMI	22.82(3.31)	22.26(20.68-24.35)	1.03	16.61(1.41)	16.82(15.77-17.77)	-0.65	20.84 (2.26)	19.96(19.17-21.87)	1.20	123.81	< .001
IQ	109.05(6.86)	110.35(104.78-113.66)	-0.85	111.55(7.79)	110.77(106.22-117.17)	-0.06	111.80(7.52)	111.18(107.46-117.38)	-0.63	4.63	.201
EDE-Q Global Score	0.60(0.83)	0.30(0.14-0.60)	2.82	3.33(1.47)	3.62(2.14-4.55)	-0.46	2.83(1.67)	2.97(1.13-4.30)	-0.25	81.41	< .001
AQ10	2.31(1.72)	2.00(1.00-3.00)	0.95	3.98(2.41)	4.00(2.00-6.00)	0.36	3.54(2.16)	3.00(2.00-5.00)	0.75	20.81	< .001

Note. AN = anorexia nervosa; AQ10 = Autism Quotient-10 item version; BMI = body mass index; EDE-Q = Eating Disorder Examination – Questionnaire version; K-W = Kruskal-Wallis; IQ = intelligence quotient; IQR = interquartile range.

Table 2

Descriptive statistics associated with the accuracy and language scores associated with the Random, Goal-Directed, and Theory of Mind trials for each participant group

	Healthy Control M(SD)	Acute AN M(SD)	Weight-Restored AN M(SD)	Kruskal test statistic	*p*-value	FWE-corrected *p*-value
Random Accuracy	1.62(0.612)	1.73(0.477)	1.70(0.434)	1.05	.591	.591
Random Language	0.53(0.610)	0.23(0.460)	0.42(0.679)	9.47	.009	.052
Goal-Directed Accuracy	1.52(0.311)	1.45(0.320)	1.57(0.308)	4.23	.121	.288
Goal-Directed Language	1.02(0.169)	0.96(0.119)	1.04(0.193)	5.91	.052	.192
Theory of Mind Accuracy	1.25(0.434)	1.12(0.325)	1.24(0.366)	4.47	.107	.288
Theory of Mind Language	1.32(0.351)	1.21(0.277)	1.37(0.338)	7.06	.029	.138

Note. AN = anorexia nervosa; FWE = familywise error.

References

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1-16.

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia, 49*(11), 2971-2984.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(2), 202-212. e207.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington D.C.

Bird, G., Catmur, C., Silani, G., Frith, C., & Frith, U. (2006). Attention does not modulate neural responses to social stimuli in autism spectrum disorders. *NeuroImage, 31*(4), 1614-1624.

Bora, E., & Köse, S. (2016). Meta-analysis of theory of mind in anorexia nervosa and bulimia nervosa: A specific impairment of cognitive perspective taking in anorexia nervosa? *International Journal of Eating Disorders, 49*(8), 739-740.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849.

Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., & Sommer, M. (2012). Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *NeuroImage, 60*(3), 1652-1661.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers, 28*(1), 1-11.

Esteban, O., Blair, R., Markiewicz, C., Berleant, S., Moodie, C., Ma, F., & Gorgolewski, K. J. (2017). poldracklab/fmriprep: 1.0. 0-rc5. *Zenedo. doi, 10*.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., . . . Snyder, M. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature methods, 16*(1), 111.

Fairburn, C. G., Cooper D Phil, D. P., Zafra, Doll D Phil, H. A., O'Connor, M. E., Bohn D Phil, D. P., Kristin, Hawker, D. M., . . . Palmer, R. L. (2009). Transdiagnostic cognitive-behavioral therapy for patients with eating disorders: a two-site trial with 60-week follow-up. *American Journal of Psychiatry, 166*(3), 311-319.

Frith, C., & Frith, U. (2005). Theory of mind. *Current biology, 15*(17), R644-R645.

Gao, Y., Rogers, J. C., Pauli, R., Clanton, R., Baker, R., Birch, P., . . . Fairchild, G. (2019). Neural correlates of theory of mind in typically-developing youth: Influence of sex, age and callous-unemotional traits. *Scientific reports, 9*(1), 1-12.

Gorgolewski, K. J. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13. 1. In: Zenodo.

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics, 5*, 13.

Hayward, H., Ashwood, K., Ahmad, J., Cooke, J., Oakley, B., Moessnang, C., . . . Loth, E. (Unpublished results). Autism in 2019: Is theory of mind a core deficit across the autism spectrum? Evidence from the Longitudinal European Autism Project.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *NeuroImage, 62*(2), 782-790.

Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2009). Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. *Social neuroscience, 4*(2), 135-152.

Kerr-Gaffney, J., Harrison, A., & Tchanturia, K. (2019). Cognitive and affective empathy in eating disorders: a systematic review and meta-analysis. *Frontiers in psychiatry, 10*, 102.

Kirkovski, M., Enticott, P. G., Hughes, M. E., Rossell, S. L., & Fitzgerald, P. B. (2016). Atypical neural activity in males but not females with autism spectrum disorder. *Journal of autism and developmental disorders, 46*(3), 954-963.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., . . . Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function, 220*(2), 587-604.

Lai, M.-C., Lombardo, M. V., Chakrabarti, B., Ruigrok, A. N., Bullmore, E. T., Suckling, J., . . . Baron-Cohen, S. (2019). Neural self-representation in autistic women and association with 'compensatory camouflaging'. *Autism, 23*(5), 1210-1223.

Leppanen, J., Sedgewick, F., Treasure, J., & Tchanturia, K. (2018). Differences in the Theory of Mind profiles of patients with anorexia nervosa and individuals on the autism spectrum: a meta-analytic review. *Neuroscience & Biobehavioral Reviews, 90*, 146-163.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders, 30*(3), 205-223.

McAdams, C. J., & Krawczyk, D. C. (2011). Impaired neural processing of social attribution in anorexia nervosa. *Psychiatry Research: Neuroimaging, 194*(1), 54-63.

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*: Nfer-Nelson.

Russell, T. A., Schmidt, U., Doherty, L., Young, V., & Tchanturia, K. (2009). Aspects of social cognition in anorexia nervosa: affective and cognitive theory of mind. *Psychiatry research, 168*(3), 181-185.

Schmidt, U., & Treasure, J. (2006). Anorexia nervosa: Valued and visible. A cognitive-interpersonal maintenance model and its implications for research and practice. *British Journal of Clinical Psychology, 45*(3), 343-366.

Schulte-Rüther, M., Mainz, V., Fink, G. R., Herpertz-Dahlmann, B., & Konrad, K. (2012). Theory of mind and the brain in anorexia nervosa: relation to treatment outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(8), 832-841. e811.

Sedgewick, F., Leppanen, J., Goh, F., Hayward, H., Happé, F., & Tchanturia, K. (2019). Similarities and differences in Theory of Mind responses of patients with anorexia nervosa with and without autistic features. *Frontiers in psychiatry, 10*, 318.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., . . . Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23*, S208-S219.

Stewart, C. S., McEwen, F. S., Konstantellou, A., Eisler, I., & Simic, M. (2017). Impact of ASD traits on treatment outcomes of eating disorders in girls. *European eating disorders review, 25*(2), 123-128.

van Veluw, S. J., & Chance, S. A. (2014). Differentiating between self and others: an ALE meta-analysis of fMRI studies of self-recognition and theory of mind. *Brain imaging and behavior, 8*(1), 24-38.

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., . . . Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage, 29*(1), 90-98.

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research, 4*(2), 149-154.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica, 67*(6), 361-370.

Zürcher, N. R., Rogier, O., Boshyan, J., Hippolyte, L., Russo, B., Gillberg, N., . . . Gillberg, C. (2013). Perception of social cues of danger in autism spectrum disorders. *PloS one, 8*(12), e81206.

Supplementary Material

Supplementary Table 1

Results of the One-sample t-test Exploratory Whole Brain Analysis for Neural Activation Associated with Increasing Complexity for the Theory of Mind Contrast

A $Z > 3.1$ cluster-forming threshold was used. We report significant clusters at the $p < .05$ threshold.

Cluster Number	Hemisphere	K	P_{FWE}	Peak Coordinates			Description
				x	y	z	
Cluster 1	Right	10,911	< .001	43.5	-46.5	-12.5	Fusiform gyrus, posterior superior temporal gyrus, and the middle temporal area
				47.5	-40.5	11.5	
				49.5	-24.5	-4.5	
				47.5	-20.5	-10.5	
				47.5	-38.5	3.5	
				41.5	-58.5	-10.5	
Cluster 2	Left	8,531	< .001	-42.5	-52.5	-14.5	Fusiform gyrus, visual association area, and the angular area
				-28.5	-98.5	-8.5	
				-24.5	-100	-2.5	
				-22.5	-102	-6.5	
				-58.5	-52.5	13.5	
				-46.5	-64.5	-14.5	
Cluster 3	Right	2,771	< .001	49.5	23.5	23.5	Dorsolateral prefrontal cortex and the inferior frontal gyrus
				39.5	7.5	27.5	
				51.5	19.5	29.5	
				35.5	9.5	29.5	
				47.5	11.5	29.5	
				57.5	25.5	27.5	
Cluster 4	Right	1,097	< .001	21.5	-60.5	25.5	Dorsal posterior cingulate cortex, ventral posterior cingulate cortex, and the precuneus
				17.5	-54.5	19.5	
				15.5	-56.5	23.5	
				11.5	-52.5	41.5	
				1.5	-62.5	37.5	
				3.5	-54.5	33.5	
Cluster 5	Right	223	< .001	17.5	-82.5	-30.5	The medial cerebellum

				23.5	-82.5	-32.5	
				13.5	-74.5	-28.5	
				25.5	-74.5	-34.5	
Cluster 6	Left	216	< .001	-20.5	-60.5	23.5	Visual association area
				-16.5	-66.5	35.5	and the precuneus
Cluster 7	Bilateral	149	< .001	-0.5	57.5	35.5	The medial prefrontal
				-8.5	61.5	35.5	cortex
				-0.5	51.5	45.5	
				3.5	49.5	39.5	
				-0.5	45.5	43.5	
				3.5	61.5	25.5	
Cluster 8	Right	120	< .001	33.5	-24.5	15.5	Primary auditory cortex
				37.5	-26.5	21.5	and the temporo-parietal
							junction
Cluster 9	Left	112	< .001	-34.5	-28.5	17.5	The temporo-parietal
							junction
Cluster 10	Right	108	< .001	3.5	17.5	67.5	The supplementary motor
				7.5	9.5	73.5	area
				1.5	11.5	63.5	
				9.5	23.5	65.5	
				11.5	5.5	73.5	
Cluster 11	Right	54	< .001	39.5	-78.5	39.5	The angular area and the
				39.5	-80.5	35.5	extrastriate cortex
				45.5	-76.5	29.5	
Cluster 12	Right	52	< .001	3.5	-26.5	65.5	The primary motor cortex
Cluster 13	Right	33	.003	17.5	-28.5	27.5	White matter
Cluster 14	Left	32	.004	-22.5	-24.5	5.5	White matter
				-20.5	-24.5	-0.5	
Cluster 15	Right	26	.014	39.5	-4.5	13.5	Primary motor cortex
Cluster 16	Right	24	.022	11.5	-26.5	-34.5	The brainstem
				11.5	-24.5	-40.5	
Cluster 17	Right	24	.022	41.5	11.5	-20.5	The temporopolar area

Supplementary Table 2

Results of the One-sample t-test Exploratory Whole Brain Analysis for Neural Activation Associated with Decreasing Complexity of the Theory of Mind Contrast

A $Z > 3.1$ cluster-forming threshold was used. We report significant clusters at the $p < .05$ threshold.

Cluster Number	Hemisphere	K	P_{FWE}	Peak Coordinates			Description
				x	y	z	
Cluster 1	Bilateral	26,613	< .001	15.5	-78.5	9.5	Right visual association area, right primary visual cortex, left visual association area, and the left primary visual cortex
				11.5	-92.5	17.5	
				-4.5	-100	17.5	
				-6.5	-94.5	13.5	
				-12.5	-86.5	7.5	
				-8.5	-98.5	11.5	
Cluster 2	Left	869	< .001	-32.5	17.5	11.5	The inferior frontal gyrus and the premotor cortex
				-40.5	11.5	7.5	
				-44.5	11.5	3.5	
				-48.5	-0.5	7.5	
				-44.5	15.5	-2.5	
				-58.5	5.5	37.5	
Cluster 3	Left	856	< .001	-28.5	33.5	27.5	Dorsolateral prefrontal cortex and the frontal eye fields
				-26.5	39.5	43.5	
				-30.5	37.5	35.5	
				-32.5	43.5	39.5	
				-30.5	33.5	47.5	
				-22.5	31.5	11.5	
Cluster 4	Right	711	< .001	31.5	17.5	11.5	Inferior frontal gyrus, premotor cortex, and posterior superior temporal gyrus
				39.5	15.5	3.5	
				45.5	3.5	7.5	
				43.5	15.5	-0.5	
				53.5	-2.5	-0.5	
				51.5	3.5	-0.5	
Cluster 5	Left	393	< .001	-14.5	-22.5	43.5	Supplementary motor area and dorsal posterior cingulate cortex
				-10.5	-20.5	47.5	
				-6.5	-22.5	49.5	

				-2.5	-32.5	37.5	
				-18.5	-34.5	41.5	
				-12.5	-16.5	43.5	
Cluster 6	Right	229	< .001	19.5	5.5	27.5	White matter
				19.5	-4.5	29.5	
				19.5	-8.5	29.5	
				19.5	15.5	23.5	
				23.5	-12.5	35.5	
				23.5	5.5	41.5	
Cluster 7	Left	177	< .001	-44.5	-66.5	-38.5	Lateral cerebellum
				-44.5	-50.5	-34.5	
				-38.5	-44.5	-32.5	
				-50.5	-60.5	-38.5	
				-44.5	-46.5	-38.5	
				-44.5	-64.5	-44.5	
Cluster 8	Left	80	< .001	-48.5	-58.5	41.5	The angular area
				-42.5	-52.5	35.5	
				-44.5	-52.5	39.5	
				-50.5	-54.5	49.5	
Cluster 9	Right	51	< .001	35.5	-44.5	-30.5	Lateral cerebellum
				37.5	-54.5	-30.5	
				33.5	-52.5	-28.5	
Cluster 10	Left	51	< .001	-12.5	39.5	23.5	Medial prefrontal cortex
				-8.5	37.5	13.5	and the dorsal anterior
				-10.5	41.5	17.5	cingulate cortex
				-12.5	41.5	7.5	
				-14.5	37.5	13.5	
Cluster 11	Right	50	< .001	5.5	53.5	-0.5	Medial prefrontal cortex
				13.5	59.5	1.5	
				11.5	55.5	1.5	
Cluster 12	Left	49	< .001	-36.5	49.5	-10.5	Frontopolar cortex
				-28.5	45.5	-10.5	
Cluster 13	Left	49	< .001	-30.5	-48.5	-48.5	Inferior cerebellum
				-26.5	-40.5	-46.5	
Cluster 14	Left	29	.007	-18.5	-52.5	-46.5	Inferior cerebellum

Cluster 15	Right	28	.009	33.5	-34.5	11.5	White matter
				29.5	-26.5	9.5	
Cluster 16	Right	23	.029	13.5	-36.5	15.5	White matter
				7.5	-34.5	11.5	
				3.5	-30.5	15.5	
Cluster 17	Left	22	.037	-40.5	-48.5	1.5	White matter
				-42.5	-40.5	-6.5	
Cluster 18	Right	22	.037	23.5	-54.5	35.5	White matter
				19.5	-52.5	41.5	
Cluster 19	Left	21	.048	43.5	-36.5	-8.5	White matter
				43.5	-42.5	-4.5	
Cluster 20	Left	21	.048	-8.5	-68.5	-34.5	Medial cerebellum
				-6.5	-72.5	-38.5	

Inclusion and Exclusion Criteria for the Study

Participants in the current AN participant group were required to meet DSM-5 criteria for AN at the point of recruitment and have a BMI less than 90% of the median BMI for age and gender or a body mass index (BMI) less than 18.5. Participants in the weight-recovered AN group must have previously been diagnosed with AN, but have had a BMI within the healthy weight range (18.5-25) during the 12-month period prior to study participation. Participants in the healthy control group were required to have no current eating disorder or history of an eating disorder. Participants in the healthy control group were also required to have a BMI within the healthy weight range (18.5-25). Exclusion criteria for the study included any neurological impairment (e.g., epilepsy), serious brain injury or learning difficulties, and MRI incompatibility (e.g., pregnancy, claustrophobia, inability to lie down flat, and any metal in or on the body which could not be removed).

Participant Medication Use

Thirty-eight women with current AN, 14 weight-restored women, 13 women in recovery from AN, and 13 healthy control women were taking medication at the time of the study. This amounted to 23% of the total sample taking some form of medication at the time of the study, less than half of whom were taking psychiatric medication (9.6% of the total sample).

With regards to psychiatric medication, 17 women with current AN were taking an antidepressant, 3 were taking an antipsychotic, 5 were taking both an antidepressant and an antipsychotic, and 1 was taking an antidepressant and benzodiazepine. Of the women recovered from AN, 5 were taking an antidepressant at the time of the study. Of the women who were weight-recovered from AN, 8 were taking an antidepressant, 1 was taking an antipsychotic, and 3 were taking an antidepressant and an antipsychotic. Two of the healthy control women

were taking antidepressants and one healthy control women was taking a stimulant for attention deficit hyperactivity disorder (ADHD) at the time of the study.

Measures

Eating Disorders Examination – Questionnaire version. The Eating Disorder Examination – Questionnaire version (EDE-Q) is a self-report measure of eating disorder psychopathology. The EDE-Q assesses the raw frequency of common eating disorder behaviours and also contains four eating disorder psychopathology subscales measuring Restraint, Eating Concern, Weight Concern, and Shape Concern. Each subscale is presented in the form of a 7-point Likert scale. For each item, participants are asked to indicate over what range of days they exhibited each component of eating disorder psychopathology, where responses are anchored from 0 ("No days") to 6 ("Every day"). Higher scores on the EDE-Q therefore indicate greater levels of eating disorder psychopathology. The EDE-Q is associated with acceptable criterion validity, with significantly different mean scores for each subscale among individuals with, versus without, a current eating disorder (18).

Hospital Anxiety and Depression Scale. The Hospital Anxiety and Depression Scale (HADS) is a 14-item self-report questionnaire assessing levels of depression and anxiety. Each item is presented on a 4-point Likert scale anchored from 0-3. The HADS yields separate anxiety and depression subscales, where higher scores on each subscale indicate greater levels of anxiety and depression, respectively. The HADS is associated with good concurrent validity, with strong positive correlations to other measures of anxiety and depression (20).

The National Adult Reading Test. The National Adult Reading Test (NART) is a measure of premorbid intellectual function in English-speaking adults. The test consists of a list of 50 written words with irregular spellings, which the participant is prompted to read aloud. The participants' ability to pronounce each word correctly tests the participants' vocabulary, which is used as a proxy measure for intelligence. Scores on the NART are converted to an

estimated intelligence quotient (IQ) score. The NART exhibits good concurrent validity, with a strong positive correlation to scores on the Wechsler Adult Intelligence Scale (22). The primary advantage of administering the NART, as opposed to the WAIS, is that it takes a fraction of the time to complete, thus reducing participant burden.

The Autism Quotient-10 item version. The Autism Quotient-10 item version (AQ-10) is a 10-item questionnaire assessing autistic symptomatology. Items are presented in the form of a 4-point Likert scale, anchored from "strongly disagree" to "strongly agree". Items are scored as either 0 or 1 depending on the direction of endorsement. Each item score is subsequently summed, such that higher scores on the AQ-10 indicate greater levels of autistic symptomatology. The AQ-10 has good sensitivity (88%) and specificity (91%) in the prediction of autism spectrum disorders (ASD) with a cut-off point of 6.0.

The Autism Diagnostic Observation Schedule. The Autism Diagnostic Observation Schedule (ADOS) is a semi-structured interview measuring autistic traits in the domains of social interaction, communication, play, and imaginative use of materials. Higher ratings within each module indicate greater levels of autistic traits. Each module has good test-retest reliability and excellent inter-rater reliability. The ADOS is associated with excellent sensitivity (82-95%) and specificity (80-100%) to detect ASD (20-22).

fMRIprep Boilerplate

Results included in this manuscript come from preprocessing performed using fMRIPrep 1.5.1rc1 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on Nipype 1.3.0-rc1 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

Anatomical data preprocessing. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

Functional data preprocessing. For each of the 1 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD

reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al. 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox and Hyde 1997, RRID:SCR_005927). The BOLD time-series, were resampled to surfaces on the following spaces: fsaverage5. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also

calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained componentsâ€ ™time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.5.2 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep documentation.

Copyright Waiver

References

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. "Machine Learning for Neuroimaging with Scikit-Learn." Frontiers in Neuroinformatics 8. https://doi.org/10.3389/fninf.2014.00014.

Avants, B.B., C.L. Epstein, M. Grossman, and J.C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain". Medical Image Analysis 12 (1): 26-41. https://doi.org/10.1016/j.media.2007.06.004.

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI". NeuroImage 37 (1): 90-101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Cox, Robert W., and James S. Hyde. 1997. "Software Tools for Analysis and Visualization of fMRI Data". NMR in Biomedicine 10 (4-5): 171-78. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.

Dale, Anders M., Bruce Fischl, and Martin I. Sereno. 1999. "Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction". NeuroImage 9 (2): 179-94. https://doi.org/10.1006/nimg.1998.0395.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. “FMRIPrep”. Software. Zenodo. https://doi.org/10.5281/zenodo.852659.

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, et al. 2018. “fMRIPrep: A Robust Preprocessing Pipeline for Functional MRI”. Nature Methods. https://doi.org/10.1038/s41592-018-0235-4.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. “Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood”. NeuroImage 47, Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. “Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python”. Frontiers in Neuroinformatics 5: 13. https://doi.org/10.3389/fninf.2011.00013.

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. “Nipype”. Software. Zenodo. https://doi.org/10.5281/zenodo.596855.

Greve, Douglas N, and Bruce Fischl. 2009. “Accurate and Robust Brain Image Alignment Using Boundary-Based Registration”. NeuroImage 48 (1): 63-72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images”. NeuroImage 17 (2): 825-41. https://doi.org/10.1006/nimg.2002.1132.

Klein, Arno, Satrajit S. Ghosh, Forrest S. Bao, Joachim Giard, Yrjö Häme, Eliezer Stavsky, Noah Lee, et al. 2017. “Mindboggling Morphometry of Human Brains”. PLOS Computational Biology 13 (2): e1005350. https://doi.org/10.1371/journal.pcbi.1005350.

Lanczos, C. 1964. “Evaluation of Noisy Data”. Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis 1 (1): 76-85. https://doi.org/10.1137/0701007.

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. “Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI”. NeuroImage 84 (Supplement C): 320-41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2013. “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data”. NeuroImage 64 (1): 240-56. https://doi.org/10.1016/j.neuroimage.2012.08.052.

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. “N4ITK: Improved N3 Bias Correction.” IEEE Transactions on Medical Imaging 29 (6): 1310-20. https://doi.org/10.1109/TMI.2010.2046908.

Zhang, Y., M. Brady, and S. Smith. 2001. “Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm”. IEEE Transactions on Medical Imaging 20 (1): 45-57. https://doi.org/10.1109/42.906424.