



ELSEVIER

Contents lists available at ScienceDirect

## Neuropsychologia

journal homepage: [www.elsevier.com/locate/neuropsychologia](http://www.elsevier.com/locate/neuropsychologia)

# Autistic adolescents show atypical activation of the brain's mentalizing system even without a prior history of mentalizing problems<sup>☆</sup>



Sarah J. White<sup>a</sup>, Uta Frith<sup>a</sup>, Julian Rellecke<sup>b</sup>, Zainab Al-Noor<sup>a</sup>, Sam J. Gilbert<sup>a,\*</sup>

<sup>a</sup> Institute of Cognitive Neuroscience, University College London, London, WC1N 3AR, UK

<sup>b</sup> Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

## ARTICLE INFO

## Article history:

Received 23 September 2013

Received in revised form

6 December 2013

Accepted 13 December 2013

Available online 19 December 2013

## Keywords:

Autism

Theory of mind

Social impairment

Development

Longitudinal

## ABSTRACT

Some autistic children pass classic Theory of Mind (ToM) tasks that others fail, but the significance of this finding is at present unclear. We identified two such groups of primary school age (labelled ToM+ and ToM-) and a matched comparison group of typically developing children (TD). Five years later we tested these participants again on a ToM test battery appropriate for adolescents and conducted an fMRI study with a story based ToM task. We also assessed autistic core symptoms at these two time points. At both times the ToM- group showed more severe social communication impairments than the ToM+ group, and while showing an improvement in mentalizing performance, they continued to show a significant impairment compared to the NT group. Two independent ROI analyses of the BOLD signal showed activation of the mentalizing network including medial prefrontal cortex, posterior cingulate and lateral temporal cortices. Strikingly, both ToM+ and ToM- groups showed very similar patterns of heightened activation in comparison with the NT group. No differences in other brain regions were apparent. Thus, autistic adolescents who do not have a history of mentalizing problems according to our ToM battery showed the same atypical neurophysiological response during mentalizing as children who did have such a history. This finding indicates that heterogeneity at the behavioural level may nevertheless map onto a similar phenotype at the neuro-cognitive level.

© 2013 The Authors. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

One of the most enduring puzzles presented by autism is the huge difference that may be observed between individuals, while at the same time there is the compelling impression of similarity at some level. The impairments in social communication and interaction, recently reaffirmed as critical for a clinical diagnosis by DSMV, may also be critical to this intuitive impression. Although the precise nature of the social impairments remains elusive, the 'Theory of mind' or 'mentalizing' hypothesis (Baron-Cohen, 1995; Baron-Cohen, Leslie, & Frith, 1985; Frith, 2012) represents a systematic attempt to explain both mild and severe social impairments in autistic individuals. However, there is a problem for this hypothesis. It has long been known that some autistic individuals can solve Theory of Mind tasks and others do not (Happé, 1995; Moran et al., 2011). The question we address here is whether these are two distinct subgroups or whether both

represent a similar neurocognitive phenotype, at the level of analysis provided by functional neuroimaging methods. We can answer the question on the basis of the presence or absence of atypical brain activation in those autistic children who have a history of being able to solve Theory of Mind tasks.

The existence of a mentalizing system in the brain is now well accepted (for reviews see Frith & Frith, 2012; Kennedy & Adolphs, 2012; Lieberman, 2007; Mitchell, 2009; Van Overwalle, 2009). There is also evidence for atypical brain activation in this system in autistic participants (Brüne & Brüne-Cohrs, 2006; Gilbert, Bird, Brindley, Frith, & Burgess, 2008; Gilbert, Meuwese, Towgood, Frith, & Burgess, 2009; Gotts et al., 2012; Kana, Libero, Hu, Deshpande, & Colburn, 2012; Lombardo et al., 2010; Marsh & Hamilton, 2011; Spengler, Bird, & Brass, 2010). However it is not known whether such atypical activation is present even in individuals who can solve Theory of Mind tasks. There may be an underlying impairment which is camouflaged at the behavioural level (Frith, 2004). Camouflage may happen when highly verbal individuals have learned to give accurate answers to Theory of Mind tests using effortful logical inferences. Thus good mentalizing performance does not necessarily imply intact intuitive mentalizing ability. Indeed problems in implicit mentalizing have been revealed in autistic adults who performed well on explicit mentalizing tasks (Begeer, Bernstein, van Wijhe, Scheeren, & Koot, 2012; Senju, Southgate, White, & Frith, 2009). On the other hand, it is possible

<sup>☆</sup>This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

\* Correspondence to: Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK. Tel.: +44 20 7679 5498, +44 20 3317 6770; fax: +44 20 7813 2835.

E-mail address: [sam.gilbert@ucl.ac.uk](mailto:sam.gilbert@ucl.ac.uk) (S.J. Gilbert).

that a subgroup of autistic individuals have no problems in understanding intentions and beliefs, or that they can fully overcome such problems. This should be evident not only at the behavioural but also the neural level. If so, this would suggest a distinct neurophysiologically defined phenotype.

To investigate this issue the present study took advantage of an existing population of autistic as well as neurotypical adolescents who had been extensively tested in childhood (White, Hill, Happé, & Frith, 2009). The sample of adolescents who participated in the present fMRI study were classified on the basis of their performance 5 years earlier on a large mentalizing test battery: Thus one subgroup consisted of those who had shown mentalizing performance as good as that of neurotypical (NT) children (ToM+), and another group consisted of those who had shown the more familiar pattern of mentalizing impairment (ToM−). We administered a second mentalizing test battery to find out to what extent performance changed over time. We also wanted to establish the validity of the mentalizing task performance. It would be pointless to classify subgroups on the performance of tests that were neither reliable over time nor valid in relation to their real world symptoms. Therefore we investigated whether ToM+ children had milder core symptoms on diagnostic tests, both in childhood and in adolescence.

## 2. Methods

### 2.1. Participants

Ethical approval for the study was received from the UCL Research Ethics Committee and consent was obtained from the parents of all participants prior to inclusion in the study. The majority of individuals with ASD attended mainstream schools and all had IQs within the normal range (full scale IQ greater than 85).

The participants were aged 11–17 years. They were a self-selected subset of those who previously took part in a study by White et al. (2009) at time 1 (T1) when aged 7–12 years. The original sample included 45 children with autism spectrum disorder (ASD), diagnosed independently by a qualified clinician. Of these, 29 were willing to be involved in the study 5 years later at time 2 (T2), but 7 of these children were either unable to tolerate the scanning environment or their data were unusable due to movement in the scanner. The final sample of 22 adolescents with ASD was split into ToM− and ToM+ groups of 11 each. A further group of 11 typically developing adolescents was also recruited from the original sample. The three groups were comparable in age ( $F(2,30)=1.56, p=.23$ ), gender ( $\chi^2(2)=.57, p=.75$ ), verbal ( $F(2,30)=.90, p=.42$ ) and performance IQ ( $F(2,30)=2.44, p=.10$ ) (see Table 1).

For the assessment of core symptoms the developmental, dimensional and diagnostic interview (3Di; Skuse et al., 2004) was used at T1, and the autism diagnostic observation schedule (ADOS-G; Lord et al., 2000) at T2. The 3Di measure is similar to the autism diagnostic interview (ADI-R; Lord, Rutter, & Le Couteur, 1994) with which it correlates highly ( $r$  for each area of the triad of autistic features ranges from .53 to .64; Skuse et al., 2004).

### 2.2. T1 ToM battery

This battery included both the ToM Strange Stories and a false belief ToM battery. The original T1 cohort included 27 NT children (White et al., 2009) and, on the basis of this larger NT group's performance, individual variance on both sets of stories was calculated, independent of age and IQ. This was done by entering raw data from these 27 NT children as the dependent variable in a regression, with T1 age, verbal, and performance IQ as predictors, and collecting the residuals. The same regression equation was then applied to the 22 children with ASD taking part in this study and all scores were converted to z-scores in relation to the larger NT group's means and standard deviations. The average of the ToM Strange Stories and the false belief ToM battery z-scores was then calculated to provide an overall measure of T1 ToM test performance for each child. This method was necessary to provide an individual estimate of ToM ability independent of age and IQ on which the children with ASD could be divided, so as to avoid the ToM− group being populated with younger and lower-IQ individuals. All participants in the resulting ToM− group had T1 ToM z-scores lower than  $-2.5$  in comparison to the T2 NT group.

### 2.3. T2 ToM battery

Of the 5 tests given at T2, two had been included in the T1 false belief battery and found then to be most discriminating between the groups, and three tasks

**Table 1**  
Behavioural data.

	NT	ToM−	ToM+
<b>Background data</b>			
N (M:F)	11 (10:1)	11 (9:2)	11 (10:1)
T2 age (years)	14.3 (1.6)	13.8 (1.1)	13.1 (1.9)
T1 WISC III Verbal IQ	118 (16)	110 (16)	111 (14)
T1 WISC III Performance IQ	103 (9)	103 (9)	95 (10)
T1 Clinical diagnosis	–	3 Autism/6 AS 2 ASD	8 AS 3 ASD
<b>Assessment of core symptoms</b>			
T1 3Di social *** <sup>a</sup>	3.4 (2.4)	14.2 (4.6)	10.3 (3.9)
Communication *** <sup>b</sup>	3.5 (2.0)	15.3 (4.3)	13.5 (3.3)
Repetitive behaviour ** <sup>b</sup>	2 (.4)	5.4 (2.8)	3.9 (2.9)
T2 ADOS social * <sup>c</sup>	–	7.1 (4.2)	3.5 (2.8)
Communication	–	3.0 (1.7)	2.0 (.8)
Repetitive behaviour	–	3.0 (2.2)	1.8 (1.3)
<b>Performance on mentalizing tasks</b>			
T1 ToM battery (z-score)*** <sup>d</sup>	.0 (1.0)	−4.5 (1.4)	−.6 (0.8)
T2 ToM battery (z-score) <sup>e</sup>	.0 (1.0)	−1.3 (1.4)	−.7 (1.3)
T2 scanner ToM Stories (%)** <sup>c</sup>	88 (10)	64 (22)	71 (17)
T2 scanner Non-ToM Stories (%)	85 (12)	75 (24)	77 (16)

Values are given as mean with standard deviation in brackets.

T1=time 1, T2=time 2.

ToM=theory of mind.

NT=neurotypical; ASD=autism spectrum disorder; AS=Asperger syndrome.

3Di=developmental, dimensional, and diagnostic interview.

ADOS=autism diagnostic observation schedule.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

<sup>a</sup> ToM− ≠ ToM+ ≠ NT.

<sup>b</sup> ToM− = ToM+ ≠ NT.

<sup>c</sup> ToM− ≠ ToM+.

<sup>d</sup> ToM− ≠ ToM+=NT.

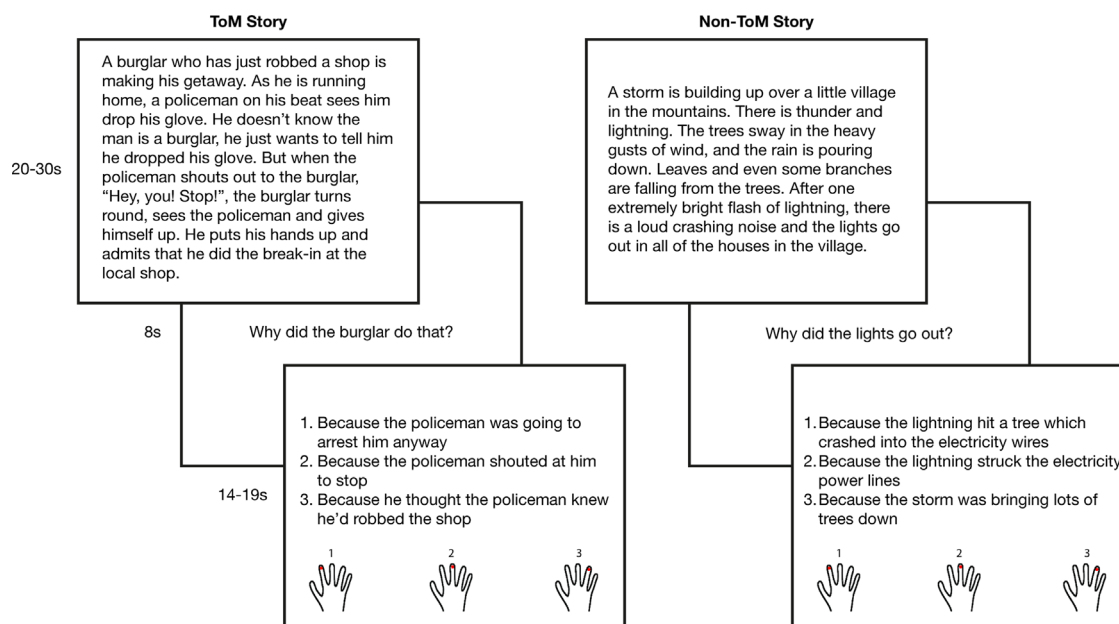
<sup>e</sup> ToM− ≠ NT.

were new to the participants. The two T1 tasks were 1st order false belief tasks: a test of real versus apparent emotion (Wellman & Liu, 2004), where a character wanted to create a false belief in others, and an interpretational false belief task (different picture to T1; Luckett, Powell, Messer, Thornton, & Schulz, 2002). The three new tasks were 2nd order false belief tasks: the coat story (Bowler, 1992), a homework story modelled on the icecream van story (Baron-Cohen, 1989), and a double bluff burglar story involving 2nd order deception (Happé, 1994). These were scored out of a total of 10. In all cases, the participant was asked to predict the knowledge of, behaviour of or emotion felt by another character on the basis of their mental state; this answer was marked as a pass or fail (1 or 0). The participant was also asked to justify why they had made that prediction; this was marked as a correct mental state justification (e.g., "because he doesn't know that she knows that they're out of stock"), a correct non-mental state justification (e.g., "because that's where he had said he was going to go"), or an incorrect justification (e.g., "because he needed to get a new coat"; 1, 0.5, or 0, respectively). This additional scoring aimed to check for false positive responses when the participant was achieving the correct answer by guessing, and also was expected to increase the variation in responses with the aim of avoiding ceiling effects. Control questions were administered (memory and reality questions plus prompt questions during the longer stories), which all individuals were required to pass, to check for comprehension of the scenario.

### 2.4. T2 scanner task

Four different types of stories were administered: mental, human, animal and nature; for the present study only mental (ToM) and nature (non-ToM) were compared as the human and animal stories have previously been found to rely to a degree on mentalizing ability (White et al., 2009). The story text and preceding question were identical to those given in White et al. (2009) but the participant was required to select an answer to this question from a choice of three possible responses. Participants listened to each story, question and response phase presented binaurally through headphones, as well as simultaneously following the text visually on a computer screen in order to aid comprehension (see Fig. 1).

The stories were presented during four blocks; each block involved 8 stories, 2 of each type, presented in a pseudorandom order, ensuring that two stories of the same type were not presented consecutively. Each story-question-response series lasted 58 s in total; the story phase was presented first (taking between 20 and



**Fig. 1.** Schematic illustration of the task administered during fMRI scanning, showing an example ToM story (correct answer: 3) and an example non-ToM story (correct answer: 1).

30 s, followed by a 1 s interval), proceeded by the question (taking 8 s) and then response options (taking between 6 and 11 s for the options to be read out, plus an additional 8 s), concluding with an interval of variable duration (as the story length was also variable). Participants selected from the multiple choice options by pressing keys on a button box with their index, middle and ring fingers on their right hand.

### 2.5. Scanning procedure

Participants were familiarised with the tasks during a practice session lasting approximately 10 min, immediately before the scanning session. A 1.5T Siemens Avanto system was used to acquire both T1-weighted structural images and T2\*-weighted echoplanar (EPI) images [64 × 64; 3 mm × 3 mm pixels; echo time (TE), 40 ms] with BOLD contrast. Each volume was comprised of 48 axial slices (3 mm thick), oriented approximately parallel to the AC-PC plane. Functional scans were acquired during four sessions, each comprising 156 volumes (lasting ~8 min). Volumes were acquired continuously with an effective repetition time (TR) of 3 s per volume. The first four volumes in each session were discarded to allow for T1 equilibration effects. Following the functional scans, a 6 min T1-weighted structural scan was performed.

### 2.6. Data analysis

The fMRI data were analyzed using SPM8 software (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). The volumes were realigned, corrected for different slice acquisition times, normalized into 3 mm cubic voxels using the Montreal Neurological Institute reference brain and 4th-degree B-spline interpolation, and smoothed with an isotropic 8-mm full-width half-maximum Gaussian kernel. The volumes acquired during the four sessions were treated as separate time series. For each series, the variance in the BOLD signal was decomposed with a set of regressors in a general linear model. Separate regressors coded for the story, question, and response phases of each story, along with an additional regressor to represent the time of motor response, separately for each of the four story types, to yield a total of 16 regressors per session. Regressors for the three phases were generated by convolving a boxcar, corresponding to the duration of each phase, with a canonical haemodynamic response function. The regressors representing motor responses (which, unlike the previous three regressors, were not entered into subsequent statistical analyses) were generated by convolving a delta function at the time of response production with a canonical haemodynamic response function. These regressors, together with the regressors representing residual movement-related artifacts and the mean over scans, comprised the full model for each session. The data and model were high-pass filtered to a cutoff of 1/128 Hz.

Parameter estimates for each regressor were calculated from the least mean squares fit of the model to the data. Effects of interest were assessed in random effects analyses using *t*-tests on contrast images generated from subject-specific analyses. Contrasts were thresholded at  $P < 0.001$  uncorrected for multiple comparisons,

with an extent threshold calculated by SPM8 to yield a family-wise error corrected probability of  $p < .05$ .

### 2.7. Assessment of head motion

In order to examine whether the groups may have differed in head motion, we calculated a mean framewise displacement (FD) measure, in the manner suggested by Power, Barnes, Snyder, Schlaggar, and Petersen (2012), (p. 2144). This yields, for each participant, a single value indicating mean head displacement from one volume to the next.

### 2.8. Functional connectivity analysis

In a further set of analyses we investigated functional connectivity between four regions of interest (ROIs) defined from the BOLD activation data. We extracted four timecourses of activation from the preprocessed images, by averaging across all voxels belonging to each ROI in turn. This was done separately for each session of each participant's data, applying a hi-pass filter identical to the one used for the BOLD activation analysis. Scans were allocated to separate ToM and non-ToM datasets when they belonged to one or the other condition, after allowing for a delay of two TRs (6 s) to account for the rise of the haemodynamic signal. Fisher-transformed correlation coefficients were then calculated for the correlations between the signal in each pair of ROIs, separately for the ToM and non-ToM conditions. These coefficients were then averaged across sessions. In the main analysis results were averaged across all pairs of ROIs; a follow-up analysis investigated those pairs of ROIs involving medial prefrontal cortex (i.e. frontal-posterior connectivity) and the other pairs (i.e. posterior-posterior connectivity) separately. This provides a simple approach for obtaining a measure of functional connectivity between a set of predefined ROIs; alternative approaches such as PPI (Friston et al., 1997) are more appropriate for investigating functional connectivity between a single seed region and the whole brain volume.

## 3. Results

### 3.1. Assessment of core symptoms in childhood and adolescence

Both 3Di and ADOS measures, used at the two different points in time, indicated milder social impairments in the ToM+ than the ToM- group (3Di:  $t(20)=2.13$ ,  $p=.046$ ,  $d=.91$ ; ADOS:  $t(20)=2.39$ ,  $p=.027$ ,  $d=1.07$ ). There was a marginally significant difference in communication impairment measured with the ADOS ( $t(14.1)=1.80$ ,  $p=.094$ ,  $d=.77$ ) but not the 3Di ( $t(20)=1.13$ ,  $p=.271$ ,  $d=.50$ ). Repetitive behaviour did not differ significantly between groups

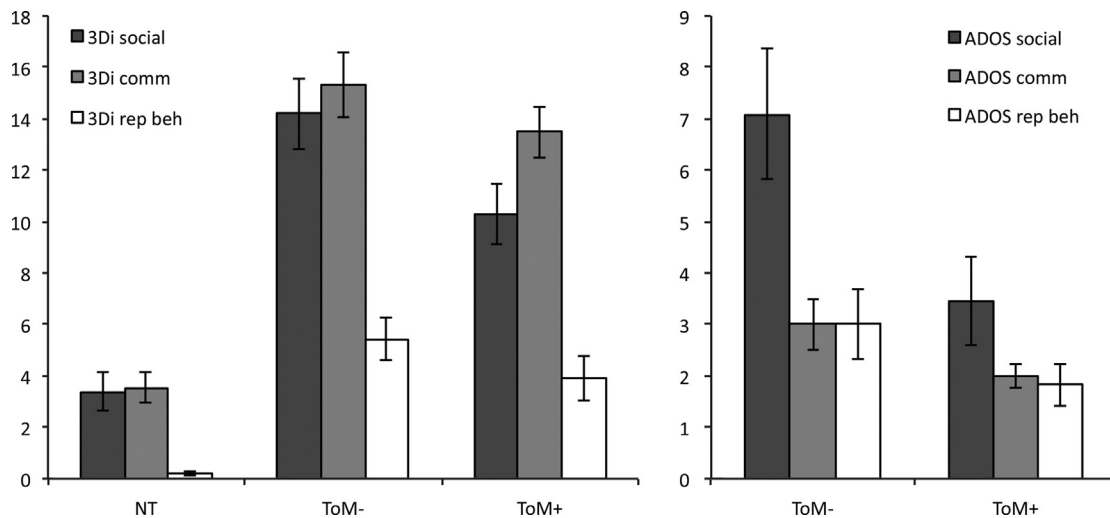


Fig. 2. Assessment of core symptoms. Error bars indicate standard error of the mean.

on either measure (3Di:  $t(20)=1.23$ ,  $p=.23$ ,  $d=.55$ ; ADOS:  $t(20)=1.56$ ,  $p=.14$ ,  $d=.70$ ) (see Table 1B and Fig. 2).

In adolescence 8 of the ToM+ and three of the ToM- participants did not obtain a score above the cut-off point for diagnosis of an ASD on the ADOS; this differentiated the two groups ( $\chi^2=4.7$ ,  $p=.030$ ). While the proportion of clinically diagnosed milder or higher-IQ cases that meet ADOS criteria can lie between 38% and 54% (Baird et al., 2006; Kamp-Becker et al., 2013), it is consistent with recent studies that suggest that in high functioning individuals the core symptoms of autism can diminish over time to the extent that they are no longer detectable (Fein et al., 2013; Gotham, Pickles, & Lord, 2012).

Supporting the milder presentation of the ToM+ group, more children in the ToM- group had received autism-specific preschool intervention (8 out of 11) than children in the ToM+ group (1 out of 11) ( $\chi^2=9.2$ ,  $p=.002$ ). However, the two ASD groups did not differ significantly on age at diagnosis ( $t(20)=1.52$ ); the ToM+ cases were no less likely to have been identified at a young age.

### 3.2. Changes in mentalizing test performance from childhood to adolescence

By definition when comparing the 3 groups' performance at T1 on a ToM test battery ( $F(2,30)=52.2$ ,  $p<.001$ ,  $\eta^2=.78$ ), the ToM- and ToM+ groups differed hugely ( $p<.001$ ) with the ToM- group performing very poorly indeed, while the NT and ToM+ groups did not differ ( $p=.206$ ) (see Table 1C and Fig. 3).

At T2, another ToM battery was used, which differentiated the groups less dramatically than before ( $F(2,30)=3.3$ ,  $p=.052$ ,  $\eta^2=.18$ ). However, the ToM- group still performed the mentalizing tasks significantly less accurately than the NT group ( $p=.016$ ). The ToM+ group performance lay between these two groups, not differing significantly from either ( $p>.16$ ).

During fMRI scanning, when participants answered questions about mental causality (ToM stories) or about physical causality (non-ToM stories), there was a main effect of Group on accuracy ( $F(2,30)=3.5$ ,  $p=.043$ ,  $\eta^2=.19$ ). Both ASD groups had generally lower accuracy than NT children across the two story types (ToM- vs NT:  $F(1,20)=6.1$ ,  $p=.023$ ,  $\eta^2=.23$ ; ToM+ vs NT:  $F(1,20)=5.9$ ,  $p=.025$ ,  $\eta^2=.23$ ). In the ToM- group, this effect was exacerbated compared to the NT group for stories requiring mentalizing (group  $\times$  story-type interaction:  $F(1,20)=4.3$ ,  $p=.052$ ,  $\eta^2=.18$ ). This effect was non-significant when comparing the ToM+ and NT groups ( $F(1,20)=1.8$ ,  $p=.19$ ,  $\eta^2=.08$ ). Indeed, the ToM- group performed worse on ToM stories than non-ToM stories ( $F(1,10)=4.8$ ,  $p=.053$ ,  $\eta^2=.33$ )

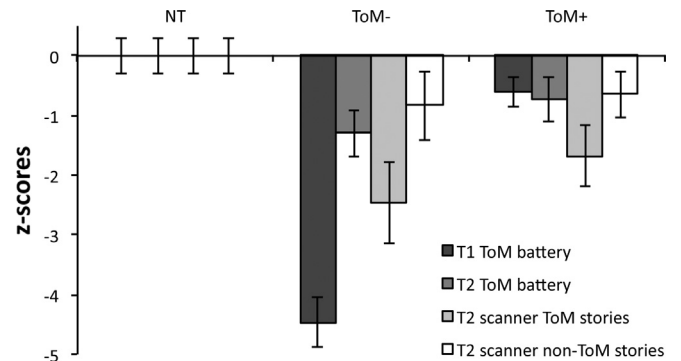


Fig. 3. Performance on mentalizing tasks. Error bars indicate standard error of the mean.

whereas there was no significant difference between accuracy on the two types of story in the ToM+ group ( $F(1,20)=1.7$ ,  $p=.22$ ,  $\eta^2=.15$ ). Thus only the ToM- group showed evidence of a mentalizing-specific impairment, relative to the NT group. Mean response times did not differ according to condition or group; nor was there a significant group  $\times$  condition interaction ( $F(1,30)<1.8$ ,  $p>.19$ ,  $\eta^2<.08$ ).

Across the different tasks, a correlation was found in the ASD group as a whole between accuracy in the T1 ToM battery and the T2 ToM stories ( $r=.48$ ,  $p=.023$ ; vs T2 non-ToM stories  $r=.33$ ,  $p=.138$ ). This remarkably high correlation between tests administered five years apart indicates the stability of the measures and a hint of specificity. This hint of specificity gains strength from the partial correlation between T1 ToM battery and T2 ToM stories ( $r=.38$ ,  $p=.094$ ), controlling for more general story comprehension (T2 non-ToM stories).

### 3.3. Brain activation patterns in the mentalizing system in adolescence

We used a region of interest (ROI) approach to compare the three groups. First, we defined the mentalizing network by performing a contrast of ToM versus non-ToM stories, collapsing over the three groups and the three story phases (presentation of the story, the question, and the response options). This revealed four regions of activation using a whole-brain family-wise-error-corrected threshold: medial prefrontal cortex, posterior cingulate, and bilaterally the temporo-parietal junction/temporal poles (Table 2; Fig. 4A). Mean signal across all voxels was then extracted



for the ToM versus non-ToM contrast in each of the four ROIs and each of the three story phases for each participant. These data were used for direct comparisons between groups.

The hypothesis testing data were independent of the data used to define the ROIs, under the null hypothesis (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). This was possible because the group comparisons involved orthogonal contrasts to the ones used to define the ROIs, and furthermore, we used a balanced design with equal numbers of participants in each of the three groups. This unbiased approach maximises our statistical power to detect group differences within the mentalizing network: Comparisons were performed within

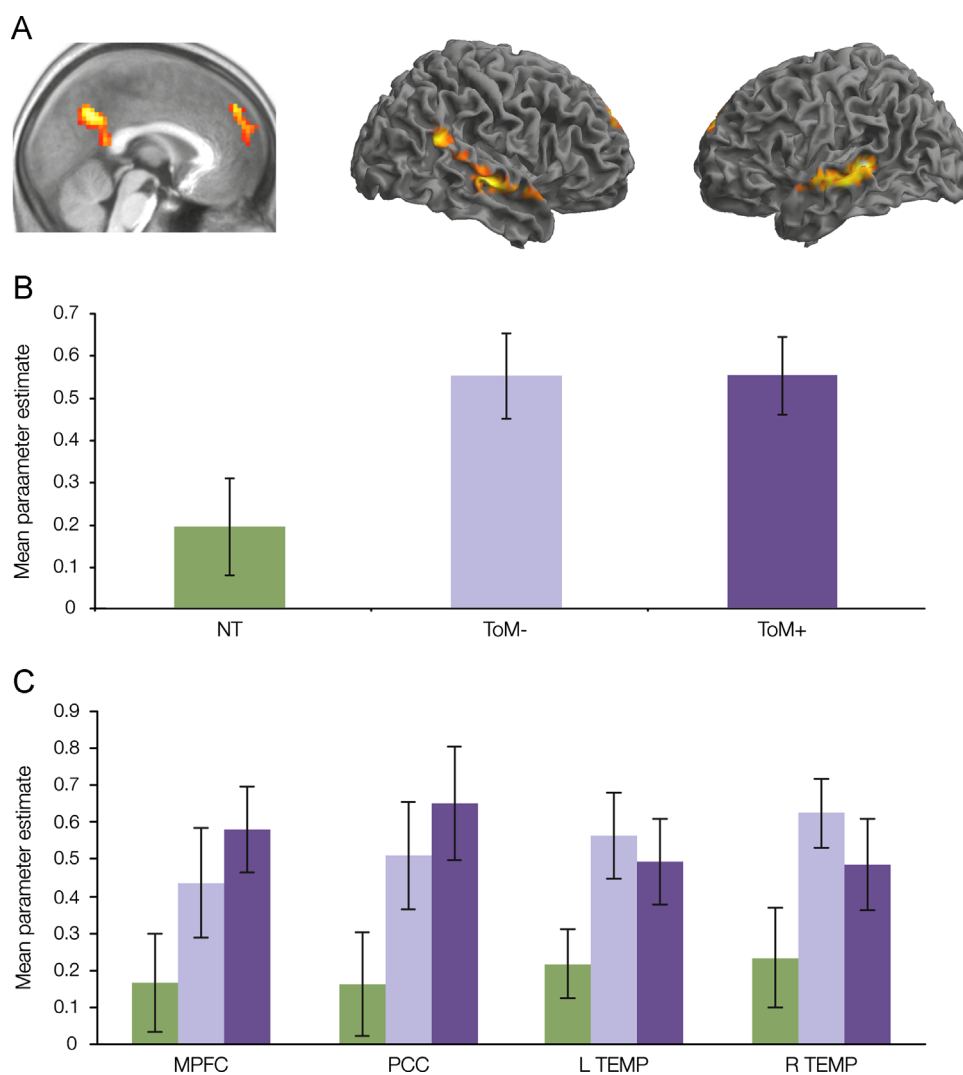
a limited number of orthogonally-defined regions of interest rather than requiring exploratory whole-brain-corrected analyses.

A 3 (group)  $\times$  3 (story phase)  $\times$  4 (ROI) ANOVA showed a main effect of group ( $F(2,30)=3.8$ ,  $p=.034$ ,  $\eta^2=.20$ ), which did not interact with any other factor ( $p > .37$ ). Note that the BOLD data entered into this analysis is already a subtraction between the two conditions, so this is equivalent to a group  $\times$  condition interaction. Follow-up tests revealed that both the ToM+ ( $F(1,20)=5.3$ ,  $p=.033$ ,  $\eta^2=.21$ ) and the ToM- ( $F(1,20)=5.4$ ,  $p=.030$ ,  $\eta^2=.21$ ) groups showed significantly greater mentalizing-related BOLD signal change than the NT group, but there was no difference between the two ASD groups ( $F(1,20)=.02$ ,  $p=.89$ ,  $\eta^2=.001$ ) (see Fig. 4B). A whole-brain analysis comparing the ToM- and ToM+ groups failed to reveal any regions of significant signal change. To illustrate mentalizing-related signal change in each of the three groups, unthresholded statistical maps for the ToM versus non-ToM comparison are presented in Fig. 5.

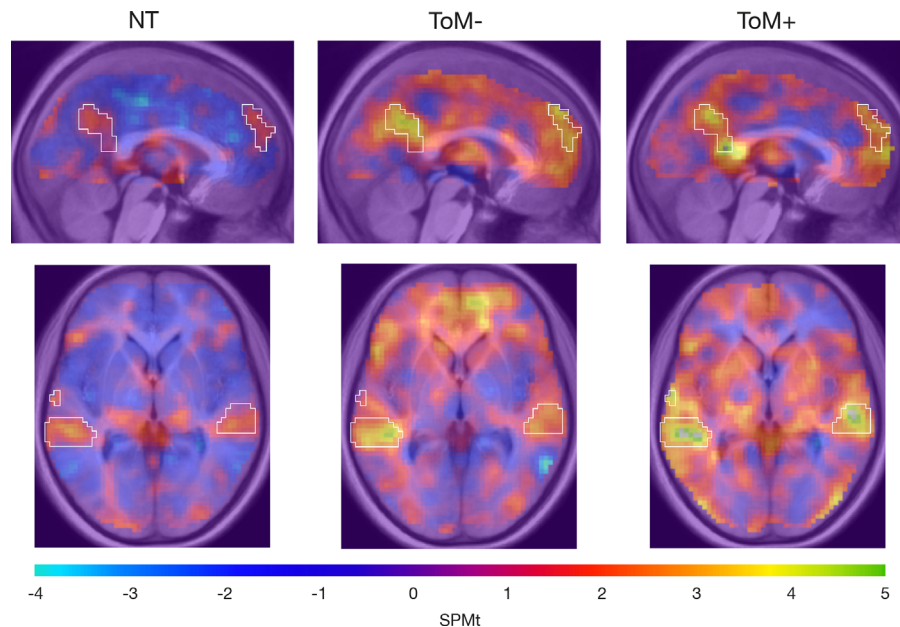
To summarize, the brain activation patterns in the mentalizing system were remarkably similar for the ToM+ and ToM- groups and differed from the NT group. However these results have to be considered in the light of some particular limitations of our study, namely the small sample size imposed by the longitudinal design.

**Table 2**  
Regions showing significantly greater activity for the ToM than the non-ToM stories, collapsed across all 3 groups (BA=Brodmann area).

Region	BA	Co-ordinate	$Z_{max}$	N voxels
Medial prefrontal cortex	9/10	12, 56, 31	4.20	173
Posterior cingulate cortex	23	3, -55, 31	4.74	155
Left lateral temporal cortex	21/22/38	-45, -37, 4	5.51	400
Right lateral temporal cortex	21/22/38	57, -25, -2	4.97	519



**Fig. 4.** Mentalizing network and signal change relating to mentalizing. Panel A: Four regions comprising the mentalizing network, illustrated on a sagittal slice of the mean structural image ( $x=0$ ) and SPM8 renderings on the right and left lateral surfaces. Panel B: Mean signal change, collapsed across the four regions and three story phases, in each of the three groups. Panel C: Mean signal change in each of the four regions of interest. Error bars indicate standard error of the mean.



**Fig. 5.** Unthresholded statistical parametric maps, in each of the groups, showing signal change associated with mentalizing. The white outlines indicate regions of interest, which were defined by comparing ToM vs non-ToM stories collapsed across the three groups. Results are illustrated on sagittal ( $x=0$ ) and axial ( $z=0$ ) slices of the mean structural scan.

#### 4. Follow-up analyses of fMRI results

##### 4.1. Effects of head motion

Analysis of the mean framewise displacement measure (FD; Power et al., 2012) showed that, compared with the TD group (FD:.13), there was a marginally significant increase in head motion in the ToM+ group (FD:.25,  $t(10.8)=1.91$ ,  $p=.08$ ,  $d=.85$ ) and a nonsignificant increase in the ToM- group (FD:.26,  $t(10.6)=1.64$ ,  $p=.13$ ,  $d=.73$ ). The ToM+ and ToM- groups did not differ ( $t(20)=.06$ ,  $p=.95$ ,  $d=.03$ ). FD did not correlate significantly with mentalizing-related BOLD signal change in any of the three groups considered individually (ToM-:  $r=-.29$ ,  $p=.39$ ; ToM+:  $r=.10$ ,  $p=.77$ ; TD:  $r=-.46$ ,  $p=.15$ ), nor across the whole sample ( $r=.24$ ,  $p=.18$ ). Thus, although there was limited evidence for increased head motion in the ToM groups compared with TD participants, this did not relate to mentalizing-related BOLD signal change.

##### 4.2. Consistency of results using ROIs derived from an independent study

The foregoing analysis defined ROIs by collapsing across all participants. However it should be noted that the majority of participants in our sample belonged to one of the two ASD groups. Consequently it is uncertain whether results would have been similar if the mentalizing network had been defined from neurotypical participants only. In order to address this issue we conducted an additional analysis using ROIs defined from a prior independent study. We based these ROIs on data reported by Burnett, Bird, Moll, Frith, and Blakemore (2009), seeing as their study investigated brain activity elicited by representation of mental states in (neurotypical) adolescent participants, using verbal stories as stimuli. Thus their study was similar to ours both in terms of participant demographics and stimulus materials. In their adolescent group, Burnett et al. report four regions of activation: medial prefrontal cortex ( $-10$ ,  $52$ ,  $18$ ); precuneus/posterior cingulate ( $-4$ ,  $-56$ ,  $28$ ), left temporal cortex ( $-38$ ,  $-66$ ,  $42$ ) and right temporal cortex ( $44$ ,  $-48$ ,  $28$ ). For each of

these four regions, we defined ROIs using a 10 mm radius sphere centered on the peak co-ordinate. Collapsing across our three groups, BOLD signal was significantly greater for ToM than non-ToM stories in three of the four ROIs ( $F(1,30)>9.9$ ;  $p<.004$ ;  $\eta^2>.24$ ); however, the left temporal ROI did not show any significant signal change, either collapsed across the three groups ( $F(1,30)=.02$ ;  $p=.90$ ;  $\eta^2<.01$ ), or in any of the three groups individually ( $F(1,30)<1.5$ ;  $p>.26$ ;  $\eta^2<.13$ ). Consequently, this ROI was dropped from further analysis. The three ROIs that did respond to the ToM versus non-ToM contrast were examined in a 3 (Group)  $\times$  3 (Phase)  $\times$  3 (ROI) ANOVA, as in the earlier analysis. This showed a marginally significant effect of Group ( $F(2,30)=2.9$ ;  $p=.07$ ;  $\eta^2=.16$ ). Follow-up tests showed that both the ToM- group ( $F(1,20)=3.1$ ;  $p=.09$ ;  $\eta^2=.13$ ) and the ToM+ group ( $F(1,20)=5.2$ ;  $p=.03$ ;  $\eta^2=.21$ ) showed elevated BOLD signal, in comparison with the TD group. The two ASD groups did not differ from each other ( $F(1,20)=.48$ ;  $p=.50$ ;  $\eta^2=.02$ ). None of these effects differed between regions ( $p>.5$ ). Thus, results were similar, even when ROIs were defined from an independent study of neurotypical participants.

##### 4.3. Did participants with ASD show an elevated BOLD signal across all contrasts?

Our fMRI results might be explained by some factor that leads to a globally elevated BOLD signal in participants with ASD, for example due to physiological reasons, regardless of the contrast or the brain regions examined. However, further analysis showed that this was not the case. An identical analysis to the foregoing analysis of ToM versus non-ToM stories was conducted. However, for this analysis the contrast of non-mentalizing stories versus rest (i.e. the inter-trial interval) was investigated; in all other respects the analysis was identical to the one presented in Fig. 4. There were four regions of activation, collapsed across the three groups: (1) bilateral occipital cortex (peak:  $24$ ,  $-97$ ,  $1$ ;  $Z_{\max}=6.78$ ; 394 voxels); (2) right lateral temporal cortex (peak:  $51$ ,  $-31$ ,  $-2$ ;  $Z_{\max}=5.52$ ; 584 voxels); (3) left lateral temporal cortex (peak:  $-63$ ,  $-4$ ,  $-5$ ;  $Z_{\max}=6.87$ ; 355 voxels); (4) right posterior frontal cortex (peak:  $39$ ,  $5$ ,  $49$ ;  $Z_{\max}=4.01$ ; 82 voxels). Results were

analysed as above in a 3 (group) × 3 (story phase) × 4 (ROI) ANOVA. In contrast to the mentalizing-related analysis, this analysis revealed no significant effect of group (mean parameter estimates: NT = .50, SE = .10; ToM- = .60, SE = .10; ToM+ = .35, SE = .09;  $F(2,30) = 1.7$ ,  $p = .20$ ,  $\eta^2 = .10$ ). Thus it was not the case that signal change was elevated in the two autism groups across all statistical contrasts, as might be expected if the results shown in Fig. 4 resulted from a global factor that differed between groups.

#### 4.4. Was there sufficient power to detect differences between ToM+ and ToM- groups?

Calculation of effect sizes (Cohen's  $d$ ) indicated large effects when comparing BOLD signal change in each of the ASD groups with the NT group (ToM-:  $d = 1.04$ ; ToM+:  $d = 1.03$ ). By contrast, the comparison of ToM+ versus ToM- yielded a  $d$  of just 0.06. There were also large behavioural differences between the ToM+ and ToM- groups in the ADOS and 3Di social scales ( $d = 1.01$  and 0.91, respectively). Power calculations (Faul, Erdfelder, Lang, & Buchner, 2007) indicate that had there been correspondingly large effects in the BOLD signal comparison between ToM+ and ToM-, our study would have power ranging from 66 to 78% to detect at least a marginally significant effect, whereas there was in fact no hint of an effect ( $p = .89$ ). However, it should be noted that post-hoc power analyses of this type are likely to inflate effect sizes (Button et al., 2013).

#### 4.5. Would a Bayesian analysis give similar results?

Rouder, Speckman, Sun, Morey, and Iverson (2009) suggest an alternative approach for determining whether a dataset can be considered to provide positive support in favour of, or against, a null hypothesis, by calculating Bayes Factors (Kass & Raftery, 1995). Applying the JZS Bayes Factor method suggested by Rouder et al. to our ToM+ versus ToM- BOLD comparison (with default scale factor 1.0) yields a Bayes Factor of 3.3, i.e. positive support for the null hypothesis, where a factor in excess of 3.2 is conventionally considered to provide "substantial" evidence (Kass & Raftery, 1995).

#### 4.6. Were there differences in functional connectivity between groups?

In a final analysis we considered whether the groups may have differed in functional connectivity between the four ROIs, even in the absence of overall signal change differences. Functional connectivity was calculated in each group, separately for the ToM and non-ToM stories, averaged across each pair of ROIs. Connectivity was significantly greater during ToM than non-ToM stories ( $F(1,30) = 8.3$ ,  $p = .007$ ,  $\eta^2 = .22$ ). However, there was no main effect of group nor group × condition interaction ( $F(2,30) < .95$ ,  $p > .39$ ,  $\eta^2 < .06$ ). The measures of functional connectivity did not correlate significantly with head motion (i.e. mean framewise displacement), either within each group separately or across the whole

sample ( $|r| < .57$ ;  $p > .069$ ). These results are shown in Table 3. Similar results held even when investigating frontal-posterior or posterior-posterior connectivity alone (effect of condition:  $F(1,30) > 5.4$ ;  $p < .03$ ;  $\eta^2 > .15$ ; effect of group/group × condition interaction:  $F(2,30) < 2.4$ ;  $p > .11$ ;  $\eta^2 < .14$ ). Thus, although our measure was sensitive enough to detect an enhancement of functional connectivity during ToM versus non-ToM stories, this measure did not distinguish the three groups.

## 5. Discussion

All the individuals in our sample had been clinically diagnosed as autistic in early childhood. When they were first tested with mentalizing tasks they were between 7 and 12 years old. Their performance on these tasks varied enormously and allowed us to divide them into two groups (ToM+ and ToM-). The differences between the groups, who were of similar IQ and socioeconomic background, were reflected in differences in symptom severity. Thus, poor mentalizing performance was associated with a more severe pattern of social and communication behaviours. This was the case in childhood, when parents were interviewed, and was still the case in adolescence when the participants were assessed in the lab through observation with the ADOS. The relationship with different assessment instruments at different points in time are in line with the idea that mentalizing difficulties underlie both performance on specific tests and core symptoms that the diagnostic instruments address.

Our longitudinal design allows us to say something about the persistence of mentalizing difficulties over time, at least for the ToM- group. Here it is likely that even early in life more severe difficulties were evident, since there was greater use of preschool intervention than in the ToM+ group. Nevertheless, mentalizing test performance improved markedly from childhood to adolescence. Despite this improvement, which brought the ToM- group more in line with the ToM+ group, they remained significantly impaired relative to the NT group. Furthermore, at the neurophysiological level both autistic groups showed equally atypical activation. We found as clear an answer as possible to the question of whether the ToM+ group did or did not differ from the ToM- group at the level of neurophysiology at adolescence: ToM+ and ToM- groups showed similarly atypical BOLD responses in the mentalizing system and both differed from the NT group. We therefore suggest that there is a neurophysiological abnormality that persists despite improvements over time, despite individual differences in performance, and is present even in mildly impaired ASD adolescents.

Both ASD groups showed *greater* activation of the mentalizing network than NT participants. This is consistent with other recent studies showing over-activation of mentalizing-related brain regions in ASD, both in mentalizing (Gilbert et al., 2009; Mason, Williams, Kana, Minshew, & Just, 2008) and non-mentalizing (Dichter, Felder, & Bodfish, 2009; Gilbert et al., 2008) tasks. However, other previous studies have reported under-activation of the mentalizing network (Castelli, Frith, Happé, & Frith, 2002; Happé et al., 1996; Kana et al., 2012; Lombardo, Chakrabarti, Bullmore, & Baron-Cohen, 2011; Silani et al., 2008; Wang, Lee, Sigman, & Dapretto, 2007) or no differences (Dufour et al., 2013). Thus, it appears that atypical mentalizing-related brain activity in ASD can take the form of both under- and over-activation, likely as a result of task-specific factors (Koster-Hale, Saxe, Dungan, & Young, 2013). Koster-Hale et al. (2013) suggest that one task-specific factor potentially influencing under- versus over-activation is the use of implicit versus explicit tasks. In an implicit task, where participants are not explicitly instructed to engage mentalizing processes, it is possible that participants with ASD fail

**Table 3**  
Functional connectivity within the mentalizing network (mean Fisher-transformed correlation coefficient) in each group.

	Mean (SD)		
	NT	ToM-	ToM+
Non-ToM stories	.527 (.107)	.570 (.179)	.567 (.190)
ToM stories	.646 (.206)	.637 (.132)	.601 (.168)



to engage these processes and therefore underactivate the mentalizing network. An example of this type of task might be the animations task used by Castelli et al. (2002), where participants are asked to describe the movement of animated shapes. Neurotypical individuals typically interpret such animations using mental-state concepts more appropriately than autistic participants (White, Coniston, Rogers, & Frith, 2011). In Castelli et al.'s (2002) study autistic participants under-activated the mentalizing system. By contrast, in explicit tasks, participants with autism may expend more effort to compensate for mentalizing difficulties, hence over-activating the mentalizing network. The present study would be an example of an explicit task: the mentalizing stories are followed by the explicit question to explain a character's behaviour in terms of mental state attribution.

A second (but related) interpretation derives from consideration of capacity limits in mentalizing. When mentalizing demands are low, it is possible that autistic participants will engage mentalizing processes to a greater degree than neurotypicals in an attempt to compensate for their difficulties with the task. (Whether this compensatory activation actually improves performance is another matter). By contrast, for very highly demanding tasks that exceed capacity limitations, autistic participants may no longer engage in mentalizing at all, leading to reduced BOLD signal compared with neurotypicals. Thus it should not necessarily be surprising if abnormalities of the mentalizing system in autism reveal themselves as increased activation in some studies and decreased activation in others.

A third possible explanation of our finding of enhanced mentalizing-related activation in ASD is that we studied adolescent participants, in contrast to the adult samples more prevalent in previous research. Studies of typically developing adolescents have revealed increased MPFC activation compared with adult participants (Blakemore, 2008). If mentalizing development is substantially delayed in autism (Happé, 1995), this delay may be recapitulated in neural response, leading to increases rather than decreases in mentalizing-related brain activity during adolescence, or no difference (see Uddin, Supekar, & Menon, 2013, for a similar argument in relation to the effect of development on atypical functional connectivity).

We are aware of the limitations of our study, in particular, the power to detect modest effect sizes. However, large effects, sufficient to yield significant results, were obtained when comparing the ToM+ and ToM− groups on two measures of social impairment, as well as when comparing BOLD signal change in each group against the NT group. This indicates that the mentalizing battery used at T1 to separate the two groups of ASD participants was sufficiently reliable to yield significant differences even at T2. The absence of any significant BOLD signal differences between ToM+ and ToM− groups cannot therefore be considered inevitable given our sample size. Power analyses suggested that the power of our study, while far from ideal, was not negligible. Furthermore, a Bayesian analysis suggested positive evidence in favour of the null hypothesis when comparing BOLD signal in the two groups, rather than merely an inability to exclude it. Thus, we contend that the present null result is noteworthy, albeit preliminary.

Our study cannot tell us what caused the better social adaptation of the ToM+ group. At this point we have no ROIs outside the standard mentalizing network to guide the search for a physiological difference that might support improved social abilities. What led this group towards better compensatory learning will have to be addressed by future research. We suggest that future studies use both implicit and explicit measures of mentalizing ability, using techniques such as eye gaze tracking. We also note that larger group sizes would allow more sensitive investigations of brain regions outside the mentalizing network. Furthermore,

alternative analytic techniques such as multivariate approaches (e.g. Gilbert et al., 2009) might shed light on individual differences in mentalizing ability and the severity of core symptoms. Our analysis of functional connectivity did not yield any significant group effects, and recent studies have produced mixed results (Müller et al., 2011; Tyszka, Kennedy, Paul, & Adolphs, 2013). Nevertheless, this approach should also be considered.

We are cautiously optimistic that despite the observed heterogeneity in autism, there is a similar neurocognitive impairment that underlies the core symptoms. Thus, mentalizing-related anomalies in brain function may map onto a consistent phenotype that underlies impairments in reciprocal social communication. This lends credence to clinical intuition that a specific impairment can unite very heterogeneous cases, hitherto not captured by diagnostic, behavioural or biological measures. If confirmed in future studies, a neuro-cognitive phenotype associated with mentalizing impairment may be useful for elucidating the genetic and neurobiological mechanisms behind autism by uniting otherwise seemingly heterogeneous cases.

## Acknowledgements

SJG is supported by a Royal Society University Research Fellowship. SJW is supported by a British Academy Postdoctoral Fellowship. We are grateful to staff at BUCNI, and to the participants in this study.

## References

- Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., et al. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *Lancet*, *368*, 210–215.
- Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*, 285–297.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Learning development and conceptual change. MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46.
- Begeer, S., Bernstein, D. M., van Wijhe, J., Scheeren, A. M., & Koot, H. M. (2012). A continuous false belief task reveals egocentric biases in children and adolescents with autism spectrum disorders. *Autism*, *16*, 357–366.
- Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*, 267–277.
- Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *33*, 877–893.
- Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—Evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience and Biobehavioral Reviews*, *30*, 437–455.
- Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, *21*, 1736–1750.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(365–76).
- 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*, 1839–1849.
- Dichter, G. S., Felder, J. N., & Bodfish, J. W. (2009). Autism is characterized by dorsal anterior cingulate hyperactivation during false belief tasks in a large sample of adults with and without autism. *PLoS One*, *8*, e75468.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.
- Fein, D., Barton, M., Eigsti, I. -M., Kelley, E., Naigles, L., Schultz, R. T., et al. (2013). Similar brain activation during false belief tasks in a large sample of individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, *54*, 195–205.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*, 218–229.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287–313.



- Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45, 672–686.
- Frith, U. (2012). Why we need cognitive explanations of autism. *Quarterly Journal of Experimental Psychology*, 1–38.
- Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, 46, 2281–2291.
- Gilbert, S. J., Meuwese, J. D. I., Towgood, K. J., Frith, C. D., & Burgess, P. W. (2009). Abnormal functional specialization within medial prefrontal cortex in high-functioning autism: A multi-voxel similarity analysis. *Brain*, 132, 869–878.
- Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, 130, e1278–e1284.
- Gotts, S. J., Simmons, W. K., Milbury, L. A., Wallace, G. L., Cox, R. W., & Martin, A. (2012). Fractionation of social brain circuits in autism spectrum disorders. *Brain*, 135, 2711–2725.
- Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129–154.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843–855.
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). "Theory of mind" in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport*, 8, 197–201.
- Kamp-Becker, I., Ghahreman, M., Heinzel-Gutenbrunner, M., Peters, M., Remschmidt, H., & Becker, K. (2013). Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism*, 17, 87–102.
- Kana, R. K., Libero, L. E., Hu, C. P., Deshpande, H. D., & Colburn, J. S. (2012). Functional brain networks and white matter underlying theory-of-mind in Autism. *Social Cognitive and Affective Neuroscience*
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
- Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, 16, 559–572.
- Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 5648–5653.
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, 12, 535–540.
- Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, 58, 259–289.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2011). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *NeuroImage*, 56, 1832–1838.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., et al. (2010). Atypical neural self-representation in autism. *Brain*, 133, 611–624.
- Lord, C., Risi, S., Lambrecht, L. E. H. C., Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A., et al. (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, 205–223.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview—revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Lockett, T., Powell, S. D., Messer, D. J., Thornton, M. E., & Schulz, J. (2002). Do children with autism who pass false belief tasks understand the mind as active interpreter? *Journal of Autism and Developmental Disorders*, 32, 127–140.
- Marsh, L. E., & Hamilton, A. F. D. C. (2011). Dissociation of mirroring and mentalising systems in autism. *NeuroImage*, 56, 1511–1519.
- Mason, R. A., Williams, D. L., Kana, R. K., Minshew, N., & Just, M. A. (2008). Theory of Mind disruption and recruitment of the right hemisphere during narrative comprehension in autism. *Neuropsychologia*, 46, 269–280.
- Mitchell, J. P. (2009). Social psychology as a natural kind. *Trends in Cognitive Sciences*, 13, 246–251.
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., et al. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2688–2692.
- Müller, R.-A., Shih, P., Keehn, B., Deyoe, J. R., Leyden, K. M., & Shukla, D. K. (2011). Underconnected, but how? A survey of functional connectivity MRI studies in autism spectrum disorders. *Cerebral Cortex*, 21, 2233–2243.
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59, 2142–2154.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, 325, 883–885.
- Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: An fMRI study. *Social Neuroscience*, 3, 97–112.
- Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., et al. (2004). The developmental, dimensional and diagnostic interview (3di): A novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 548–558.
- Spengler, S., Bird, G., & Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry*, 68, 1148–1155.
- Tyszka, J. M., Kennedy, D. P., Paul, L. K., & Adolphs, R. (2013). Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral Cortex*.
- Uddin, L. Q., Supekar, K., & Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience*, 7, 458.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30, 829–858.
- Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2007). Reading affect in the face and voice: Neural correlates of interpreting communicative intent in children and adolescents with autism spectrum disorders. *Archives of General Psychiatry*, 64, 698–708.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75, 523–541.
- White, S., 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, 149–154.
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, 80, 1097–1117.