What explains variability in brain regions associated with Theory of Mind in a large sample of neurotypical adults and adults with ASD?

Nicholas Dufour[1], Elizabeth Redcay[2], Liane Young[3], Penelope L Mavros[1], M Joseph Moran[4], Christina Triantafyllou[1,5], John Gabrieli[1,5], and Rebecca Saxe[1,5]

1. Department of Brain and Cognitive Sciences, MIT
2. Department of Psychology, University of Maryland
3. Department of Psychology, Boston College
4. Psychology Department, Harvard University
5. McGovern Institute for Brain Research, MIT

Abstract

Theory of mind ('ToM') tasks elicit highly reliable neural activity across individuals and experimental paradigms. We compared activity in a very large sample of neurotypical ('NT', N=477) individuals, and a group of high functioning individuals with autism spectrum disorders ('ASD', n=27), using both region of interest ('ROI') and whole-brain analyses. Although ToM activity showed significant and reliable individual differences, these differences were not explained by participant gender or age, or most experimental parameters. Furthermore, there were no differences between ASD and NT individuals. These results imply that the social cognitive impairments typical of ASD can occur without gross changes in the size or response magnitude of ToM brain regions.

Keywords: Theory of mind; ASD; fMRI; TPJ; PC; precuneus; MPFC; DMPFC; MMPFC

Introduction

Theory of Mind ('ToM') is the capacity to represent the mental states of others (Premack & Woodruff, 1978). Individuals with autism spectrum disorders (ASD) appear to have particular difficulty with aspects of ToM. In particular, children with ASD are disproportionately delayed on tasks that tap inferences about other people's beliefs (Baron-Cohen, 1989). The neural mechanism of this deficit remains unknown. However, in neurotypical (NT) adults and children, fMRI studies reveal a remarkable reliable group of brain regions recruited during ToM tasks. These regions include the left and right temporo-parietal junction (RTPJ and LTPJ), right anterior superior temporal sulcus (rSTS), the medial precuneus (PC), and the medial prefrontal cortex (MPFC) (U. Frith & Frith, 2003). Thus, a tempting hypothesis is that dysfunction of the brain regions typically implicated in ToM is responsible for the social cognitive impairments observed in ASD.

Previous attempts to characterize the function of ToM brain regions in adults with ASD have yielded conflicting results. Some studies suggest that ToM regions are hypoactive (i.e., produce a smaller or less selective response, (Kennedy & Courchesne, 2008; Lombardo, Chakrabarti, Bullmore), while other studies find no difference between ASD and NT individuals (Gilbert, Bird, Brindley, Frith, & Burgess, 2008), and still others find the opposite pattern, hyperactivation, in ASD (Dichter, Felder, & Bodfish, 2009).

One explanation of these conflicting results may be that sample sizes are small, and individual variability is large. Small samples of individuals with ASD are problematic because individuals with ASD may be highly heterogeneous in their neural responses (e.g., Hasson et al., 2009). Small samples of NT participants are equally problematic, because they allow for calculation of only the mean response, not the distribution. Understanding the distribution is critical if neural measures are to be useful in a clinical setting. For most clinical applications, it is more important to be able to describe the neural activity pattern of each specific individual, relative to typical and atypical distributions. For example, using fMRI to help diagnose ASD would require comparing each individual to the typical distribution.

In order to measure the distribution of responses in ToM brain regions of NT participants, we aggregated data collected over 5 years from 477 NT participants. This large sample allowed us to investigate variability in ToM region responses, and measure any difference between NT participants and adults with ASD, with unusually high sensitivity. The main goal of the current paper is therefore to compare the response in these regions in a large sample of NT participants and a moderate sample of high-functioning adults with ASD. In order to do so, we also (i) identify and remove variance in the measured response, associated with basic experimental parameters such as the stimulus modality, number of stimuli, or experimental task, and (ii) test whether the response of ToM regions is related to basic demographic factors that may be relevant for ASD, including gender, age, and IQ.

Methods

Typical Participants: Data were analyzed from 477 NT participants (M=25.2 years, range: 18-69 years; 179 male). IQ was measured in 60 of these participants (IQ 84-141, M=117.5, SD=12.4). Participants provided informed consent, in accordance with the guidelines of the MIT Committee on the Use of Human Experimental Subjects (COUHES), and were compensated approximately $30 per hour for their time.

ASD Participants: 27 individuals with a clinical diagnosis of ASD (M=33.9yrs, range 18-66yrs; 20 male) were

included, having volunteered to participate in one of two (Moran et al., 2011; Redcay et al., 2012) previous studies. The Autism Diagnostic Observation Schedule (ADOS) was administered to 23 of the 27 ASD participants (ADOS communication score M=3.2, SD=1.3; ADOS social score M=5.8, SD=1.8). For 24 of the ASD participants, IQ measures were obtained by the Kaufman Brief Intelligence Test (IQ 69-141, M=116.3, SD=16.8). For direct NT vs. ASD comparison, a set of 24 NT participants (collectively termed 'matched') were chosen based on pairwise similarity with the ASD participants on IQ, age, and gender (age 20-54, M=29.9 years, SD=8.8 years; IQ 84-141, M=116.3, SD=14.5; 19 male).

fMRI Tasks All participants were presented with verbal narratives in English that described a character and his/her mental states (Mental condition) or described physical objects and events (Control condition). The stimuli were presented either visually as text on a screen, or aurally through headphones. After reading or hearing the narrative, participants performed one of 4 tasks. These tasks correspond to the functional localizers used in (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2010; Kliemann, Young, Scholz, & Saxe, 2008; L. Young & Saxe, 2008; L. Young & Saxe, 2009; L. Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; L. Young, Nichols, & Saxe, 2010; L. Young, Scholz, & Saxe, 2011) and unpublished data.

fMRI Methods: Participants were scanned on a 3T Siemens scanner at the Martinos Imaging Center at the McGovern Institute for Brain Research at the Massachusetts Institute of Technology (n=468) or at the Massachusetts General Hospital (n=36). NT participants were scanned between 2006 and 2011. ASD participants were scanned between 2007 and 2010. Matched NTs were scanned between 2007 and 2010. Functional data were acquired using single echo gradient echo echo-planar-imaging with voxel size 3.125 x 3.125 x 4.000 mm (TE=30 ms, flip angle=90°, TR either 2.5 (n=36) or 2 secs (n=468)). Participants were scanned on either a 12-cahnnel or a 32-channel receive coil, both Siemens products. Data were analyzed using SPM2 or SPM8 (http://www.fil.ion.ucl.ac.uk) and in-house software. The data were realigned to account for motion, smoothed with a 5 mm Gaussian kernel and normalized to a standard template in Montreal Neurological Institute space.

ROI Analyses: Six functional ROIs (ROIs) from the ToM network were defined in individual participants, using the contrast Mental>Control, consistent with previous literature (e.g. (U. Frith & Frith, 2003; Saxe & Kanwisher, 2003)): RTPJ, LTPJ, PC, dorsal and middle MPFC (DMPFC and MMPFC) and rSTS.

To identify individually-defined functional ROIs, initial "hypothesis spaces" were defined as the 9mm radius sphere centered about local maxima for each region, in the group random effects analysis performed on all 477 NT participants (see figure 1). Each participant's contrast image (Mental>Control) was masked with the six hypothesis spaces; all voxels contiguous with the peak voxel and significant at $p < 0.001$, within a 9mm radius, were defined as the ROI. From each ROI three parameters were extracted: the peak voxel t-value, the size of the ROI (number of voxels included), and the mean T. The presence or absence of an ROI was used as a fourth parameter. The reliability of ROI parameters was assessed by split-half analysis. Contrast images were derived from even versus odd runs in each participant. ROIs were picked using a minimum cluster size of 10 and a significance level of $p < 0.05$. The correlation of the ROI even and odd parameter values was measured across participants.

Every subject for whom we had complete demographic and experimental data was then included in a multivariate general linear model (GLM). The resulting model was a seven-column (age, gender, group, modality, coil, number of stimuli, and the mean term) predictor matrix and included data from 383 participants. For the binary statistic that indicated whether or not the ROI of interest was identified in a given subject, the GLM presumed a binomial distribution and a logit linker function. The GLM used a normal distribution otherwise. Continuous regressors were mean-centered prior to regression. Correction for multiple comparisons was performed with Bonnferroni correction, across all predictors and all dependent measures, within each ROI. In total there were six predictors for the four ROI parameters, a total of 24 comparisons per ROI; thus effects were taken to be significant if $p < 0.0021$. Any relationship significant at $p < 0.01$ is discussed as a 'trend.'

An identical procedure was conducted for the matched group, except that coil and modality did not vary within and thus were omitted. IQ was added to the predictor matrix, resulting in a total of 20 comparisons per ROI, and a significance threshold of $p < 0.0025$. Any relationship found to have a significance $0.01 < p < 0.0025$ is discussed as a trend.

Whole-brain analyses: Whole-brain analyses were conducted for the contrast of interest (Mental>Control), in order to identify effects on the ToM brain regions. To correct for comparisons, nonparametric whole-brain analysis was performed using SnPM (http://www.sph.umich.edu/ni-stat/SnPM/). Each test used 3mm variance smoothing and 5,000 permutations, with no global normalization, grand mean scaling, or threshold masking. The corrected *p*-value for filtering was 0.05, with a threshold of 3, and a voxel-cluster combining theta value of 0.5. Permutations were repeated for each predictor of interest; all demographic and experimental predictor variables were included as nuisance regressors using modified SnPM plugins. Because (to foreshadow our results) we find a *lack* of significant differences between ASD and NT participants, we also examined the results using a substantially more lenient threshold: regions were considered significant if composed of a contiguous cluster of at least ten voxels at a t-value of 3 or greater, as this

corresponds to $p < 0.001$ (uncorrected). This more lenient threshold is consequently a more stringent test of the hypothesis that there are no differences between the groups.

Results

ROI results

Six functional ROIs (ROIs) from the ToM network were defined in individual participants, using the contrast Mental>Control, consistent with previous literature (U. Frith & Frith, 2003; Saxe & Kanwisher, 2003): RTPJ (in 414/504, or 82.1%), LTPJ (77.2%), PC (84.7%), DMPFC (60.1%), MMPFC (64.7%) and rSTS (65.5%).

The goal of this project is to explain individual differences in the size and magnitude of brain regions involved in ToM. Before testing individual differences, however, it was critical to determine that (i) there was variability in these measures, and (ii) the differences between individuals on these measures are reliable (i.e. that inter-individual differences do not simply reflect noise in the measurement). All ROI parameters showed reasonable variability. The standard deviation of the peak T-value ranged between 1 and 2, and the standard deviation of ROI size (in voxels) ranged from 60 to 90 voxels. In order to test whether this variability reflects stable individual differences, we compared the ROI measurements within individuals. ROIs were picked independently from even and odd runs in the 235 participants from whom we had more than three runs of data. RTPJ was identified in 72% of participants, LTPJ in 66%, PC in 75%, DMPFC in 55%, MMPFC in 53%, and rSTS in 56%. Correlations between the even and odd parameter values (mass, x coordinate, etc.) and across subjects had an average Pearson's *r*-value of 0.51. These correlations were all significant at $p < 0.001$, and all but two at $p < 0.0001$. Thus, the ROI parameters are reliable within subject, making it worthwhile to explain inter-subject variability.

Next we used multivariate general linear regression analyses to estimate whether any variance in the size or response magnitude of ToM brain regions is explained by ASD status. The first set of analyses compared all of the individuals with ASD (n=27, 23 male) to all of the NT individuals (n=439, 179 male). In these analyses, no parameter of any ROI was significantly predicted by the group membership (ASD vs. NT) of the individual ($p > .09$ for all ROIs). The ASD participants were similar to NT participants on the ROI measures considered; no ASD participant fell outside of 3 standard deviations on any measure or any ROI, and only one ASD participant fell outside 2 SDs. In a second set of analyses, we compared individuals with ASD (N=24, 19 male) to a group of matched NT individuals (N=24, 19 male). Again, we found no significant difference between groups on any ROI parameter (all $p > 0.01$). The new parameter of IQ was found to predict larger sized PC ROIs ($p = 0.0064$, $\beta=2.591\pm2.699$, +1.7 voxels/IQ point) at the level of a trend. Finally, the effect of ADOS score was considered. For this analysis, participants were restricted to those from the ASD group. None of the parameters significantly predicted any measured ROI parameter, even at the level of the trend.

The choice of coil had the largest effect. The 32-channel coil produced significantly greater peak (means: $p = 0.0004$, $\beta = 0.610\pm0.470$, 1.21 units higher in 32-channel ROIs) and mean T values (means: $p = 0.0006$, $\beta=0.309\pm0.212$, 0.62 units) in all ROIs except DMPFC and PC compared with the 12-channel coil. PC mean ($p = 0.0030$, $\beta = 0.253\pm0.260$, 0.541 units) and peak T ($p = 0.0039$, $\beta = 0.480\pm0.509$, 1.01 units) was increased in the 32-channel as well, but at the level of a trend. The 32-channel coil additionally significantly increased the size of the RTPJ ($p = 0.0001$, $\beta = 40.35\pm30.38$), and increased the probability of finding the RSTS ($p = 0.0087$, $\beta = 1.354\pm1.589$, 152% more likely) and its size ($p = 0.0030$, $\beta = 24.495\pm25.203$, 58.7 voxels larger) at the level of a trend. We also found an unexpected effect of number of stimuli: as the number of stimuli used in the experiment increased, the probability of identifying regions in the medial prefrontal cortex (MMPFC and DMPFC) decreased (means: $p=0.0073$, $\beta=-0.067\pm0.079$, ~ -2%/

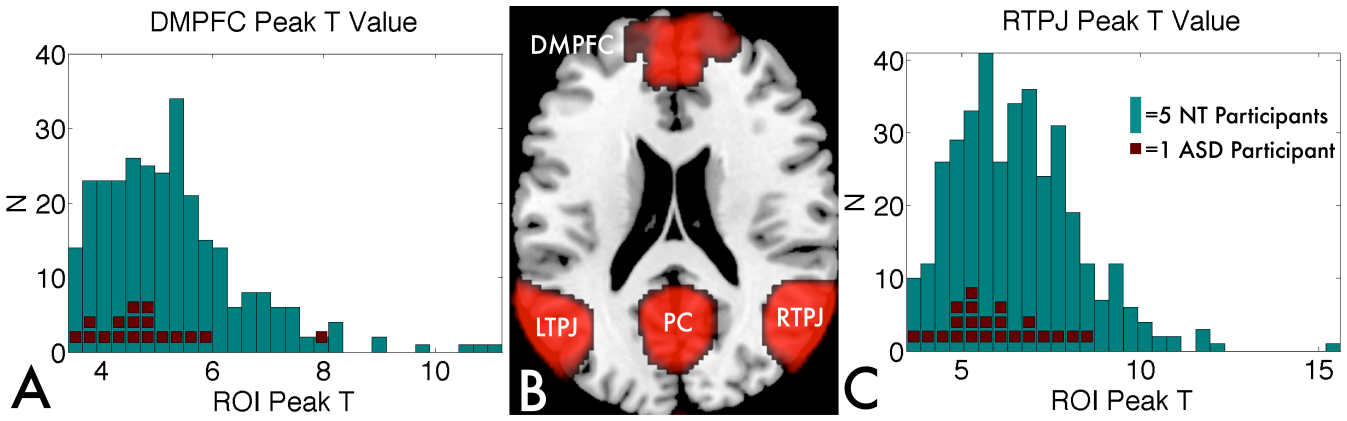

Figure 1: (A) Histogram of the DMPFC ROI peak T value for NT participants (teal) overlaid with the ASD participants (red). (B) Whole brain random effects analysis of the main effect, Mental > Control, in the full sample, corrected for multiple comparisons with permutations; axial slice shown at z = 22mm. Visible are RTPJ, LTPJ, DMPFC, and PC. (C) Histogram of the RTPJ ROI peak T value for NT participants (teal) overlaid with the ASD participants (red).

stimulus) at the level of a trend. There were no significant effects of age or gender on any parameter of any region.

In sum, ROI analyses suggest that while individuals differ reliably in the size and response magnitude of brain regions associated with ToM, these neural parameters are not affected by whether an ASD diagnosis. Experimental parameters, such as the MRI coil used, and demographic variables, such as IQ, explain some of the variance across individuals. Within the range of ADOS scores in the current sample, autism severity does not explain variance in ROI parameters, either. However, ROI analyses (and especially the three ROI parameters) provide a limited window on the brain, so to look further for differences between groups in ToM brain regions, we next conducted whole brain analyses.

Whole brain analysis results

In the whole-brain analyses, the main effect identifies brain regions significantly recruited during Mental compared to Control conditions, controlling for variance explained by any of the nuisance regressors. This analysis identified robust activation in all of the regions previous associated with ToM, including RTPJ, LTPJ, medial PC and posterior cingulate, MPFC, and anterior STS. It also identified activation in other regions, including the left superior frontal gyrus (BA8 and BA6), the left medial frontal gyrus (BA8), regions of right middle frontal gyrus (BA6, BA8, BA9), the right superior temporal gyrus (BA38) and the right inferior frontal gyrus (BA47). Also present was activity in the cingulate (BA24) and anterior cingulate (BA32), as well as the thalamus (BA24) and the right amygdala.

Next, we compared activation in the full sample of individuals with ASD vs. NT. Regions were significant if the difference between activation during mental versus control tasks was greater in one group than in the other. When these analyses were corrected for multiple comparisons using permutations, we observed no regions of significant group differences. A more lenient threshold revealed a small region in the right cingulate gyrus ([2, 14, 22], peak T 3.5, 128mm^3) with a greater condition difference in ASD than NT groups. In this region, ASD participants showed greater deactivation in the control condition, but no difference during the Mental condition. There were no regions with greater difference between conditions in the NT participants.

We also compared the ASD group to a smaller matched group. When correcting for multiple comparisons with permutations, we again failed to find any regions significant for the ASD>NT contrast. More lenient traditional thresholds also failed to reveal any significant regions. In the reverse contrast (NT > ASD), a single region was found in the right middle occipital gyrus ([36, -62, -8], peak T = 5.09, 1032mm^3) when corrected with permutations. This region was again identified using the more lenient threshold ([36, -62, -8], peak T = 5.8, 784mm^3), along with regions in the left middle temporal gyrus ([-48, 10, -44, peak T = 4.44, 304mm^3), the right middle posterior cingulate ([26, -68, 12], peak T = 4.32, 736mm^3), the left posterior lobe ([-44, -60, 38], peak T = 4.31, 488mm^3), the left cingulate gyrus ([-16, -56, 26], peak T = 4.24, 424mm^3), left inferior temporal gyrus ([-58, -28, -20], peak T = 4.24, 208mm^3), the right posterior insula ([40, -24, 12], peak T = 4.12, 304mm^3), right precentral gyrus ([32, -26, 68], peak T = 3.81, 168 mm^3), right superior temporal gyrus ([44, -60, 34], peak T = 3.77, 232mm^3), and the left posterior cingulate ([-14, -54, 14], peak T = 3.76, 104mm^3).

Discussion

The main question we sought to address in this paper was whether individuals diagnosed with ASD show differences in the magnitude or extent of activity in ToM brain regions, compared to a large sample of NT participants. To this end, we aggregated data across multiple experiments to produce a large sample of NT individuals (N=477) and a moderate sample of high functioning ASD individuals (N=27). Before directly comparing them, we tested whether neural responses to Mental stimuli were reliable within participants and variable across participants, in the NT population. They were. Next, we tested whether the magnitude of neural responses to Mental vs Control stimuli differed between groups, either in targeted regions of interest or in whole brain analyses. For the most part, these analyses identified no reliable differences between groups, especially in the previously hypothesized ToM brain regions. These results suggest that differences between these groups of participants in ToM brain regions, if they exist, are small and could not be used to diagnose ASD.

We used two complementary analysis strategies: ROI analyses focused on previous identified ToM brain regions are more sensitive, whereas whole brain analyses find group differences anywhere in the brain, and are less restricted. For both kinds of analyses, we conducted two comparisons by regression with simultaneous nuisance regressors to control for demographic and experimental variance: the ASD group vs. the whole group of NT individuals, and the ASD group vs. NT individuals matched to the ASD group on age, gender, IQ and experimental parameters. For both comparisons, we found no reliable differences between groups in the size, response magnitude, or probability of identifying above-threshold voxels, in any ToM ROI. Indeed, the ROI parameters of individuals with ASD fell squarely within the distribution of typical values, almost never straying more than 2SD from the typical means. Also, ADOS scores of the ASD participants did not predict any ROI parameter, even at the level of a trend.

In the whole brain analyses, the results of group comparisons depended on the thresholds used for correcting for multiple comparisons. Permutation-based correction, which estimates the false positive rate empirically, revealed no significant differences between the two complete groups. When we reduced the sample to just the matched NT group, we found one region, in the right middle occipital gyrus, which showed increased response to Mental than Control stimuli in the NT group, but not the ASD group. However,

since this region did not show a higher response to Mental than Control stimuli in the overall main effect analysis of all participants, and is not typically associated with any kind of social cognition, we are cautious about making strong claims based on this effect.

Because these results suggest a null result - namely, no difference between groups - we also examined the same analyses at a more lenient threshold that could reveal true differences between groups that are just below the threshold for significance. Again we found no regions more active in the full NT sample, compared to the ASD group. A small (128 mm^3) region in right cingulate gyrus appeared more active in participants with ASD at this threshold; in this region, ASD participants showed greater deactivation to the control condition than NT participants. Reducing the sample to just the matched NT participants, and using the lower threshold, produced a number of small regions showing greater activation in NT than ASD participants. However, none were in any region in the main effect analysis of Mental > Control stimuli. Thus, we could not identify any region that both (a) was reliably recruited for Mental more than Control stimuli in 477 NT individuals, and (b) showed less activity in the same contrast, in individuals with ASD.

Using a similar analysis strategy, we also found that age and gender do not affect activity in ToM brain regions; nor do the modality of the stimuli (visual vs aural) or the experimental task. Thus, although individual differences in ToM brain regions are reliable and robust, they are not explained by simple demographic or experimental variables. The absence of an effect of gender is particularly noteworthy, because the full sample contained a large number of male and female participants. Behavioral measures of ToM often reveal an advantage for female participants (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997); apparently this advantage is not due to grossly different ToM brain regions.

One significant factor was the coil used. The 32-channel coil has documented higher SNR (Triantafyllou, Polimeni, & Wald, 2011); we found that this difference translated into larger ROIs that were more likely to be detected in individual participants. Thus, our results suggest that for individually-defined ROI analyses, the increased SNR of the 32-channel coil provides a clear benefit. On the other hand, increasing the number of stimuli per condition did not have the same benefit: medial prefrontal regions were less likely to be identified, in experiments using more stimuli. This unexpected effect could reflect habituation, after more than 20 stories about characters' false beliefs.

With regard to our key null results, the current study has advantages and disadvantages. On the one hand, the large sample size provides more power and sensitivity to detect effects where they exist. In particular, although our sample of ASD individuals was only moderately large, the very large sample of NT individuals included gives us very high confidence on the true mean of the ROI parameters in NT individuals. Finding that the ASD population mean does not differ from the NT mean is thus strong evidence that these groups' data cannot be attributed to different population distributions.

However, these results cannot be interpreted as ruling out any differences in the neural mechanisms for ToM in individuals with ASD. One qualification of the current results is that the parameters measured here provide only a limited measure of a region's function. Other measures include the functional connectivity of each region and within-region spatial pattern of responses (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Haxby et al., 2001). Participants with ASD may differ in these other measures of ToM region function (Kleinhans et al., 2008). Indeed work in our lab using multi-voxel pattern analysis (MVPA) demonstrated the existence of reliable differences between ASD and NT individuals (Koster-Hale, Saxe, and Young, submitted).

Another qualification is that the ASD participants in the current sample are very high functioning. Although they meet diagnostic criteria for ASD (and have been shown to have behavioral deficits in ToM tasks in a previous study, Moran et al., 2011), these individuals are highly verbal and pass first-order false belief tasks. Thus, our results do not rule out gross differences in the ToM regions of lower-functioning individuals with ASD. On the other hand, the individuals in our sample are diagnosed with ASD because of disproportionate difficulties with social interaction and communication, and are similar to populations used in previous fMRI studies. Also, we found no evidence that within our participants, increasing ASD severity had any effect on the measured ROI parameters. So the current results imply that social cognitive impairments can occur without gross changes in the size or position of ToM brain regions. Collectively, the current results provide strong evidence that the neural differences between high functioning adults with ASD and NT participants are not due to gross changes in the magnitude of ToM brain region activity.

These results leave open a number of key questions. First, it will be key to identify the neural differences between adults with ASD and NT individuals that account for behavioral differences in ToM. One key possibility is that individuals with ASD are highly heterogeneous, so that different neural sources explain the behavioral delays in different individuals. If so, the group-average analyses used here may have limited sensitivity to detect those differences. Second, the current study focused on adults. It will be important in future research to test whether the developmental trajectory of ToM brain regions differs in children with ASD compared to NT children, even if the mature states of the system are reasonably similar. Finally, it would be useful to extend these analyses to lower-functioning individuals with ASD. Nevertheless, the implication of this study is that social-cognitive impairments can occur without large changes in the activation of ToM brain regions.

Acknowledgements

This paper is based upon work supported by the Simons Foundation, the National Science Foundation (grant 095518), the Dana Foundation, a National Science Foundation Graduate Research Fellowship (grant 0645960), and a John Merck Scholars Grant. The authors wish to acknowledge Marina Bedny, Emile Bruneau, Hyowon Gweon and Jorie Koster-Hale for collecting fMRI data.

References

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813-822.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 30*(2), 285-297.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine, 34*(4), 537-541.

Dichter, G. S., Felder, J. N., & Bodfish, J. W. (2009). Autism is characterized by dorsal anterior cingulate hyperactivation during social target detection. *Social Cognitive and Affective Neuroscience, 4*(3), 215-226.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2010). fMRI item analysis in a theory of mind task. *NeuroImage,*

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences, 358*(1431), 459-473.

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*(9), 2281-2291.

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., et al. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research, 2*(4), 220-231.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Kennedy, D. P., & Courchesne, E. (2008). Functional abnormalities of the default network during self-and other-reflection in autism. *Social Cognitive and Affective Neuroscience, 3*(2), 177-190.

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., et al. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain, 131*(4), 1000-1012.

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia, 46*(12), 2949-2957.

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,*

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, 108*(7), 2688-2692.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences, 1*(4), 515-526.

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., et al. (2012). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human Brain Mapping,*

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people:: The role of the temporo-parietal junction in. *NeuroImage, 19*(4), 1835-1842.

Triantafyllou, C., Polimeni, J. R., & Wald, L. L. (2011). Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. *NeuroImage, 55*(2), 597-606.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences, 107*(15), 6753.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology, 1*(3), 333-349.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage, 40*(4), 1912-1920.

Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience, 21*(7), 1396-1405.

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience, 6*(3), 302-315.