

Title: Individual differences in the theory of mind and superior temporal sulcus

Yuki Otsuka¹, Naoyuki Osaka¹, Takashi Ikeda¹, and Mariko Osaka²

¹ *Department of Psychology, Graduate School of Letters, Kyoto University, Kyoto 606-8501, Japan*

² *Department of Psychology, Osaka University, Suita 565-0871, Japan*

The number of text pages: 20

The number of figures: 1

The number of tables: 3

Corresponding author: Yuki Otsuka, Department of Psychology, Graduate School of Letters, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan.

Tel: 08-75-753-2478 Fax: 08-75-753-2478

E-mail address: yotsuka@bun.kyoto-u.ac.jp (Y. Otsuka).

Key words

theory of mind, individual differences, superior temporal sulcus, temporo-parietal junction,
mirror neuron system area

Abstract

We investigated the brain area with regard to individual differences in the theory of mind. Using functional magnetic resonance imaging, we examined the brain area in which signal intensity was apparently related to performance of a theory of mind task on multiple regression analysis. A significant relation was observed between performance of theory of mind task and activation in the left anterior superior temporal sulcus. We could not find such an activation in the superior temporal sulcus and the temporo-parietal junction area. The present findings provide new evidence that the anterior superior temporal sulcus might dictate individual differences in theory of mind.

Introduction

When we receive a communication from another person, we try to estimate the other person's mental state, for which theory of mind (ToM) is required. In recent years, many neuroimaging studies have investigated findings in the brain network related to ToM [1, 2, 4, 7, 15, 19]. Frith and Frith [7] proposed that the medial prefrontal cortex (mPFC), temporal pole, and posterior superior temporal sulcus (STS) support ToM-related mental tasks. In addition to these brain areas, it was shown by subsequent research that the temporo-parietal junction (TPJ) and the anterior cingulate cortex (ACC) are also concerned with ToM [2, 15]. However, several recent studies have noted a relation between ToM and the mirror neuron system (MNS), which is known to be concerned with language or imitation [1]. Hadjikhani et al. [9] examined the structure of the brain in autistic subjects and showed that the cause of autism might be a structural difference in the MNS area containing the inferior frontal gyrus (IFG), inferior parietal lobule, and STS.

While the brain network supporting ToM has been clarified, there are mainly three different theories regarding which brain area is most crucial for ToM. Amodio and Frith [2] claimed that mPFC containing ACC is especially important for ToM since they found mPFC activation during various ToM tasks, their hypothesis is supported by some imaging data [6, 12]. However, there is another theory that TPJ activation would be essential for the ability to reason regarding the other person's mental states, especially in ToM [8, 15]. Using the text of a short story, Saxe and Kanwisher [16] discovered increased activation of the bilateral TPJ when the other person's mental states had to be reasoned. Furthermore, there is a theory that the MNS area is essential in understanding the others' mind especially using empathy. Vollm et al. [20] showed different brain networks between tasks in which empathy is needed and

those in which empathy is not needed. Schulte-Ruther et al. [17] discovered activation in MNS area including IFG and STS in relation to empathy in response to an emotional facial expression.

Moreover, when considering ToM, individual differences comprise an essential factor. Although it is unquestionable that all humans have ToM, we do not always succeed in accurately estimating other people's mental states. In addition, there are some persons who excel in estimating other people's mental states, while some persons usually do not make successful estimates. It would become a major clue in searching for the brain area that is decisive for ToM to clarify the neural basis associated with individual differences in ToM. Therefore, this study examined the brain with regard to individual differences in ToM by identifying the brain area showing increased signal intensity related to performance of a ToM task on functional magnetic resonance imaging (fMRI) using multiple regression analysis. As the ToM task, we used a modified version of the task used in the study by Ferstl and von Cramon [6]. Because Ferstl and von Cramon [6] found not only mPFC activation but also activations of TPJ and the MNS area (IFG and STS). In this study, we needed to use the task in which those three areas showed activation. Among mPFC, TPJ, and the MNS area, we expected that the activation intensity of the area relevant to individual differences in ToM would show a relation to the performance of ToM task on multiple regression analysis.

Materials and methods

Participants

Twenty-four right-handed graduate and undergraduate students from Kyoto University (16 men and 8 women; mean age = 24 years, range 19-31) participated in this study.

Administered Tasks

The period of each trial was fixed to 20 s with the following time course. The start cue was initially presented for 500 ms and followed by a sentence for 5200 ms, inter-stimuli-interval of 500 ms and a second sentence for 5200 ms. Immediately after the second sentence disappeared, the cue requiring a response from the participant appeared. That cue disappeared with the participant's response (YES or NO) and a blank screen continued until the trial finished.

We required participants to perform two tasks; ToM and Control tasks. During the ToM task, participants were required to judge whether or not the second sentence was consistent with the first sentence or not based on a character's mental state. In the Control task, participants were required to judge whether or not the second sentence was consistent with the first sentence based on tense (past, present, and future). Table 1 shows examples of the sentences used in the ToM task and the Control task. We prepared 60 pairs for the ToM task and 18 pairs of sentences for the Control task. In addition, in order to confirm that the participant was reading the whole sentence in the Control task, we carried out a recognition test after the experiment. The recognition test consisted of 24 sentences; 12 sentences used in the Control task and 12 novel filler sentences.

For event-related design, we intermixed the stimuli from each task as follows. First, we created four counterbalancing lists of 60 pairs of ToM stimuli. Within each list, half of the stimuli were original pairs and the remaining were mixed pairs. The 18 Control trials were intermixed with ToM trials. Half of the stimuli were same-tense pairs and the remaining were not. The order of the trials was randomized separately in each task. The order of each task was randomized with the constraint of not more than three consecutive trials with the same type of task. The experiment consisted of 78 trials. (Table 1 about here)

fMRI data acquisition

A 1.5-T fMRI scanner (Shimazu-Marconi Magnex Eclipse) was used to acquire imaging data. Head movement was minimized using a forehead strap and soft pads under the head. For functional images, 20 images with a 6-mm thickness were acquired using the following parameters: TR, 2000 ms; TE, 48 ms; flip angle, 80°; and FOV, 256×256 mm. Anatomical images were acquired after the experiment using the following parameters: TR, 12 ms; TE, 4.5 ms; flip angle, 20°; pixel matrix, 256×256; and FOV, 25.6×25.6cm. Stimuli were generated and synchronized with the imaging sequence of the scanner using Presentation software by Neurobehavioral Systems, San Francisco, CA. Subjects viewed stimuli on a projected screen via a mirror.

fMRI data analysis

Data were analyzed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK) in MATLAB (MathWorks, Sherborn, MA). The first 6 images in the scan sequence were excluded from analysis to rule out non-equilibrium effects of magnetization, and the remaining 841 volumes of functional images were realigned to compensate for potential signal declination caused by head movement. As two participants showed head movement >1 mm during the acquisition of functional images, only images from the remaining twenty-two participants were used for analysis. After realignment, the anatomical image was coregistered to the mean functional image. Functional images were then normalized with the anatomical image and spatially smoothed using a Gaussian filter (7 mm full-width half-maximum). Task-related activity was identified with the synthetic hemodynamic response function provided by SPM. For the event-related model, we

time-locked the BOLD responses 9500 ms after the onset of the first sentence based on findings in the previous study [6]. Data were high-pass filtered with a frequency cut-off at 32 s, the duration of the task alternation period, and low-pass filtered using a hemodynamic response function. A random effects model was applied with a voxel-level threshold of $p < 0.001$ uncorrected for multiple comparisons. We applied an uncorrected criterion because we focused on specific regions in which increased activation has previously been reported during similar tasks [6]. The brain atlas by Talairach and Tournoux [18] was used to identify the anatomical regions activated.

Following identification of activation areas, percent signal changes in regions of interest (ROI) were obtained by MarsBaR [3]. We then performed a multiple regression analysis, using percent signal change of each ROI, in the STATISTICA statistical software (version 06J, by StatSoft, Tulsa, OK).

Results

The behavioral data demonstrated high levels of accuracy (ToM task, Mean = 93%, SD = 5.1, range 77-98; Control task, Mean = 93%, SD = 9.3, range 61-100), indicating that participants had efficiently completed the tasks. In addition, the performance of recognition test for Control task after an experiment were also sufficiently high (Mean = 84%, SD = 8.8, range 71-100). Imaging data for all subjects was therefore included in the following analysis.

Table 2 shows the main activation areas for each contrast. Based on the results of the contrast comparing ToM tasks with Control tasks, we set six brain areas as ROIs. Each ROI region and the center coordinate of spheres were set as follows: the left inferior frontal gyrus (IFG; -50, 28, -2), the bilateral anterior STS (-58, -8, -10; 54, -8, -14), the left posterior STS (-50, -32, -2) and the bilateral TPJ (-46, -62, 22; 54, -56, 16). ROIs were defined as

spheres with radii of 3 mm, only for left TPJ with radii of 5 mm. Fig.1 displays the mean signal changes in each ROI during the ToM and Control tasks.

We performed a forward stepwise multiple regression analysis of the mean accuracy of ToM task, using mean signal changes in each ROI on ToM task, from which the mean signal changes on Control task was subtracted, as predictors. The results of multiple regression analysis are displayed in Table 3. We found that the signal intensity in the left anterior STS accounted for a positive variance in the performance of ToM task. (Table 2, 3 and Fig. 1 about here)

Discussion

This study examined the brain area in which signal intensity was related with the performance of ToM task on multiple regression analysis. Our findings showed that the signal intensity of the left anterior STS was related to the performance of ToM task, indicating a possibility that the left anterior STS would be associated with individual differences in ToM. However, the signal intensity of mPFC and TPJ was not relevant to the performance of ToM task. Below, we discuss individual differences in ToM, and its relation to the brain area supporting ToM.

The anterior superior temporal sulcus

We demonstrated that individual differences in ToM were related to the anterior STS, not the posterior STS that has previously been indicated to have a strong relationship with ToM. Our data differ from previous findings related to this issue because we investigated individual differences in ToM. The relation between the anterior STS and performance of ToM task means that some participants would always show low activation of the anterior STS. This might be why the anterior STS during ToM task was underrepresented

by a general method seeking common brain areas among participants. Our theory is that the anterior STS might mediate a process involved in individual differences in ToM, whereas the posterior STS might mediate the fundamental ToM process in which there are no individual differences. The data from Lissek et al. [11] supports the theory that there is a difference between the anterior STS and posterior STS with regard to their roles in ToM. However, our data do not demonstrate any specific mental process leading to individual differences in ToM. Further research on ToM function in the anterior STS would clarify why there are individual differences when normal adults estimate the mental states of other persons.

Since anterior STS activation is often found in language comprehension [5, 10, 15] and we used verbal stimuli in this experiment, there may be an interpretation that anterior STS activation reflected individual differences in language comprehension instead of those in ToM. However, this interpretation seems to be denied by previous data that examined individual differences in working memory. Working memory is known to play an essential role in high cognition and also affects individual differences in sentence reading [14]. Since it was previously demonstrated that the dorsolateral prefrontal cortex and ACC would account for individual differences in working memory [13], not the anterior STS but the mPFC would have shown a relation with individual differences in ToM if performance of the ToM task reflected individual differences in sentence reading. Therefore, the interpretation that the anterior STS reflects individual differences in ToM seems to be appropriate.

Temporo-parietal junction and medial prefrontal cortex

Although we considered mPFC and TPJ as other possible areas in relation to individual differences in ToM, mPFC and TPJ did not show any such tendency. Especially mPFC did not show significant activation on contrast images made by subtraction of the Control task from the ToM task. This finding differs from those of Ferstl and von Cramon [6],

who found activation of mPFC during a similar task. It is supposed that such a difference would arise because they used pseudo-sentences whereas we used normal sentences for the Control task. In addition, Ferstl and von Cramon [6] reported mPFC activation with regard to coherence of two sentences. Therefore, our finding which was not able to discover significant mPFC activation without the factor of coherence seems to be consistent with Ferstl and von Cramon [6]. Furthermore, our finding that ToM task showed a relation with not mPFC but anterior STS also provides a clue to help resolve the recent problem of overlap between the language network and ToM network in the brain. ToM tasks are often accompanied by the possibility that participants might have verbalized information in their mind even when non-verbal stimuli are used. Accordingly, it has been discussed whether the use of language causes the overlap often found between the language network and ToM network [5]. However, we discovered that individual differences in ToM would be affected by anterior STS, not by mPFC which is likely to be related to individual differences in sentence reading. These findings seem to indicate overlap between the language network and ToM network rather than verbalization, however those networks are likely to function in a somewhat different way between language and ToM.

Although activation of TPJ was significant on the contrast image made by subtraction of the Control task from the ToM task in our study, multiple regression analysis did not show a relation with the performance of ToM task. Our finding that the TPJ showed higher activation during the ToM task than during the Control task is consistent with the hypothesis that the TPJ is concerned with the ability to reason the mental state of another person [8, 15]. Despite this hypothesis, it is interesting that the TPJ did not show a relation to individual differences in the ToM task. This also supports the possibility that the ability to reason the mental state of another person would be dissociated from ToM function arising in

the anterior STS. In a future study, it will be necessary to examine why the ToM function of the anterior STS would especially affect individual differences in ToM, and what the difference is between the anterior STS and TPJ in their roles related to ToM.

Conclusion

We examined the brain with regard to areas related to individual differences in ToM by clarifying the brain area in which signal intensity is related to the performance of ToM task using multiple regression analysis of fMRI findings. Our findings showed that signal intensity of the left anterior STS is related to the performance of ToM task.

Acknowledgements

This work was supported by research grants from the Japan Society for the Promotion of Science (20830041, 19203032, 16203037 to YO, NO, and MO, respectively) and in part by the Global COE Program "Revitalizing Education for Dynamic Hearts and Minds," MEXT, Japan.

References

- [1] Z.K. Agnew, K.K. Bhakoo, B.K. Puri, The human mirror system: a motor resonance theory of mind-reading, *Brain Res. Rev.* 54 (2007) 286-293.
- [2] D.M. Amodio, C.D. Frith, Meeting of minds: the medial frontal cortex and social cognition, *Nat. Rev. Neurosci.* 7 (2006) 268-277.
- [3] M. Brett, J.-L. Anton, R. Valabregue, J.-B. Poline, Region of interest analysis using an SPM toolbox, Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in *Neuroimage* 16 (2002) Abstract 497.
- [4] M. Corbetta, G. Patel, G.L. Shulman, The reorienting system of the human brain: from environment to theory of mind, *Neuron* 58 (2008) 306-324.
- [5] E.C. Ferstl, J. Neumann, C. Bogler, D.Y. von Cramon, The extended language network: A meta-analysis of neuroimaging studies on text comprehension, *Hum. Brain Mapp.* 29 (2008) 581-593.
- [6] E.C. Ferstl, D.Y. von Cramon, What does the frontomedian cortex contribute to language processing: coherence or theory of mind?, *Neuroimage* 17 (2002) 1599-1612.
- [7] U. Frith, C.D. Frith, Development and neurophysiology of mentalizing, *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 358 (2003) 459-473.
- [8] M.I. Gobbini, A.C. Koralek, R.E. Bryan, K.J. Montgomery, J.V. Haxby, Two takes on the social brain: a comparison of theory of mind tasks, *J. Cogn. Neurosci.* 19 (2007) 1803-1814.

- [9] N. Hadjikhani, R.M. Joseph, J. Snyder, H. Tager-Flusberg, Anatomical differences in the mirror neuron system and social cognition network in autism, *Cereb. Cortex* 16 (2006) 1276-1282.
- [10] G. Hein, R.T. Knight, Superior Temporal Sulcus-It's My Area: Or Is It?, *J. Cogn. Neurosci.* 20 (2008) 2125-2136.
- [11] S. Lissek, S. Peters, N. Fuchs, H. Witthaus, V. Nicolas, M. Tegenthoff, G. Juckel, M. Brune, Cooperation and deception recruit different subsets of the theory-of-mind network, *PLoS ONE* 3 (2008) e2023.
- [12] J.P. Mitchell, M.R. Banaji, C.N. Macrae, General and specific contributions of the medial prefrontal cortex to knowledge about mental states, *Neuroimage* 28 (2005) 757-762.
- [13] N. Osaka, M. Osaka, H. Kondo, M. Morishita, H. Fukuyama, H. Shibasaki, The neural basis of executive function in working memory: an fMRI study based on individual differences, *Neuroimage* 21 (2004) 623-631.
- [14] C.S. Prat, T.A. Keller, M.A. Just, Individual differences in sentence comprehension: A functional magnetic resonance imaging investigation of syntactic and lexical processing demands, *J. Cogn. Neurosci.* 19 (2007) 1950-1963.
- [15] R. Saxe, S. Carey, N. Kanwisher, Understanding other minds: linking developmental psychology and functional neuroimaging, *Annu. Rev. Psychol.* 55 (2004) 87-124.
- [16] R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind", *Neuroimage* 19 (2003) 1835-1842.
- [17] M. Schulte-Ruther, H.J. Markowitsch, G.R. Fink, M. Piefke, Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy, *J. Cogn. Neurosci.* 19 (2007)

1354-1372.

- [18] Talairach J, Tournoux P, *A Co-planar Stereotaxic Atlas of the Human Brain*, Thieme, New York, 1988.
- [19] F. Van Overwalle, Social Cognition and the Brain: A Meta-Analysis, *Hum. Brain Mapp.* 30 (2009) 829-858.
- [20] B.A. Vollm, A.N.W. Taylor, P. Richardson, R. Corcoran, J. Stirling, S. McKie, J.F.W. Deakin, R. Elliott, Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task, *Neuroimage* 29 (2006) 90-98.

Table/Figure Legends

Table 1: Sentence Examples for each task.

Table 2: Region of Activation for each Contrast.

Table 3: Summary of multiple regression analyses of the performance in Theory-of-Mind task, using signal change in ROIs as predictors.

Fig. 1: Activation areas were detected by comparing ToM tasks with Control tasks. The threshold for significant activation was $P < 0.001$ uncorrected at the cluster level. The mean signal changes during ToM and Control tasks are shown in each region of interest (ROI). The green bar indicates ToM task and the yellow bar indicates Control task.

Table 1

Sentence Examples for each task

ToM	Control
Sayuri set three alarm clocks by her bedside.	That red brick building collapsed in the quake.
The National Center Test will begin at eight tomorrow.	An electric fan continued to slowly circulate the damp air.

Table 2

Region of Activation for each Contrast

Brain region activation	BA	Coordinates			T value	Voxels
		x	y	z		
ToM						
middle frontal gyrus	R9	48	24	38	6.68	2081
inferior frontal gyrus	R45/47/44	54	26	6	8.97	
anterior cingulate cortex	R32	8	6	50	7.4	1011
middle frontal gyrus	L6	-6	14	50	8.68	
	L6	-36	10	58	6.08	4704
	L9	-48	22	28	5.92	
inferior frontal gyrus	L45	-52	24	2	10.1	
posterior STS	L21	-54	-26	-4	8.35	
anterior STS	L21	-56	-18	-6	6.9	
posterior STS	R21	48	-32	-2	7.77	710
anterior STS	R21	54	-14	-10	4.24	
inferior parietal lobule	L7	-36	-62	48	4.4	125
Control						
middle frontal gyrus	R46	36	36	22	5.16	241
inferior frontal gyrus	R45	46	20	8	7.28	1423
anterior cingulate cortex	R32	6	28	34	5.6	977
middle frontal gyrus	R6	4	14	50	7.64	
	L6	-6	10	50	8.32	
	L6	-28	4	64	5.43	143
	L46	-28	44	26	4.73	2174
inferior frontal gyrus	L45	-44	16	6	7.04	
transverse gyrus	L41	-44	-24	14	5.69	244
inferior parietal lobule	R7	38	-60	54	5.99	950
	L7	-32	-58	48	6.28	1909
ToM - Control						
inferior frontal gyrus	L47/45	-50	28	-2	5.44	145
posterior STS	L21	-50	-32	-2	6.23	259
anterior STS	L21	-56	-8	-10	7.65	
	R21	58	-12	-10	6.53	192
temporoparietal junction	L39/22	-46	-62	22	7.34	476
	R22	54	-56	16	6.31	95

Note : uncorrected P <.001

Abbreviations: BA=Brodman area; L=left; R=right

Table 3

Summary of multiple regression analyses of the performance in Theory-of-Mind task, using signal change in ROIs as predictors

	B	SE B	β	Δr^2	model R^2
L anterior STS	0.22	0.09	0.46 *	0.22	0.3
L inferior frontal gyrus	0.07	0.05	0.26		

Note: B, raw parameter estimate; SE B, standard error of estimate; β , standardized parameter estimate; Δr^2 , adjusted R square; model R^2 , accountable variance in the performance of Theory-of-Mind task for model with variable in percent signal change of ROIs. L, left; R, right.

** $p < .05$*

