Impaired self-awareness and theory of mind: An fMRI study of mentalizing in alexithymia

Yoshiya Moriguchi, Takashi Ohnishi, Richard D. Lane, Motonari Maeda, Takeyuki Mori, Kiyotaka Nemoto, Hiroshi Matsuda, and Gen Komaki

Department of Psychosomatic Research, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1, Ogawa-Higashi Cho, Kodaira City, Tokyo 187-8553, Japan

Department of Psychiatry, Psychology and Neuroscience, General Clinical Research Center, Health Sciences Center, The University of Arizona, PO Box 245110, Tucson, Arizona 85724-5110, USA

Department of Nuclear Medicine, Saitama Medical School Hospital, 38 Morohongo Moroyama-machi, Inuma-gun, Saitama 350-0495, Japan

Department of Radiology, National Center Hospital for Mental, Nervous, and Muscular Disorders, National Center of Neurology and Psychiatry, 4-1-1, Ogawa-Higashi Cho, Kodaira City, Tokyo 187-8553, Japan

Department of Nuclear Medicine, Saitama Medical School Hospital, 38 Morohongo Moroyama-machi, Inuma-gun, Saitama 350-0495, Japan

Department of Psychiatry, Psychology and Neuroscience, General Clinical Research Center, Health Sciences Center, The University of Arizona, PO Box 245110, Tucson, Arizona 85724-5110, USA

Graduate School of Art and Design, Joshihi University of Art and Design, 1900 Asamizodai, Sagamihara, Kanagawa 228-8538 Japan

Received 13 February 2006; revised 5 April 2006; accepted 10 April 2006

Alexithymic individuals have difficulty in recognizing and describing emotions in themselves. We investigated the neuronal basis of mentalizing in alexithymia to determine whether there is a common neuronal substrate associated with knowing the mental states of the self and others. Individuals high in alexithymia (n = 16) and low in alexithymia (n = 14) were selected from a pool of 310 college students using a combination of the Toronto Alexithymia Scale (TAS-20) and the Structured Interview version of the Beth Israel Questionnaire (SIBIQ). We compared the two groups on psychological measures, including ratings of mentalizing and the Interpersonal Reactivity Index (IRI), and regional brain activation using functional magnetic resonance imaging (fMRI) during a mentalizing animation task. The results for both groups showed activation in regions associated with mentalizing: medial prefrontal cortices (MPFC), temporo-parietal junctions (TPJ), and the temporal pole (TP). Alexithymics had lower mentalizing and IRI perspective-taking scores and less activation in the right MPFC. Activity in the MPFC was positively correlated with the mentalizing score and the IRI perspective-taking score. Although there were no group differences in cerebral activity in the TPJ and the TP, the activity in the right TP had a positive correlation with mentalizing and IRI personal distress scores. These results suggest that alexithymic individuals have an impairment in mentalizing associated with an inability to take the perspective of others. Thus, the skills involved in comprehending the self and others are inter-related and play an important role in emotion regulation.

© 2006 Elsevier Inc. All rights reserved.

Introduction

Alexithymia is a disturbance in both affective and cognitive functioning characterized by difficulty in describing or recognizing the emotions of the self. Alexithymia was originally described by Sifneos (1972) in patients with psychosomatic disorders. Subsequently, alexithymia was regarded as an impairment of emotion self-regulation that is found in a broad range of physical and psychiatric disorders (e.g., alcoholism, drug addiction, and post traumatic stress disorders; see Taylor et al., 1997). At the present time, alexithymia is not considered a discrete disorder but rather a personality characteristic that is expressed with variable intensity in the general population.

Although alexithymia refers to a deficit in emotional self-awareness, Bydlowski et al. (2005) reported that high alexithymic patients with eating disorders showed impairment in the ability to describe the emotional experiences of others in hypothetical situations. Thus, emotional self-awareness seems to be closely related to sensitivity to the emotions of others. Furthermore, the capacity to differentiate the emotions of the self from those of another person in a given context appears to be crucial for managing a variety of emotional states (Bydlowski et al., 2005). Lane and Schwartz (1987) noted that as the level of emotional awareness increases, the differentiation of self from other increases. In the absence of such differentiation, emotions remain global and undifferentiated, leading to a relative inability to use one’s own emotions to guide the selection of adaptive behavior.

Understanding that others have beliefs, desires, and intentions different from the self is a cognitive skill known as “Theory of mind” (ToM) or ‘mentalizing’ (Frith and Frith, 2003). Autistic spectrum disorders, including Asperger’s syndrome, are characterized by an impairment of ToM (Baron-Cohen et al., 1985, 1997).
Asperger’s syndrome is associated with high alexithymia scores (Berthoz and Hill, 2005; Frith, 2004; Hill et al., 2004). This suggests that the ability to describe the mental states of self and other is related. Alexithymia was also associated with impairment in the ability to identify emotions from facial expressions (Pandey and Mandal, 1997; Parker et al., 2005; Lane et al., 1996). Alexithymia has been related to certain psychiatric disorders characterized by a deficit in the ability to know what others are thinking and feeling or a lack of empathy, for example, schizophrenia (Cedro et al., 2001; Maggini and Raballo, 2004a,b; Stanghellini and Ricca, 1995; Todarello et al., 2005; van ‘t Wout et al., 2004), and borderline (Guttman and Laporte, 2002) and psychopathic personality disorders (Haviland et al., 2004). Thus, disorders characterized by ToM impairment are also associated with alexithymia.

It would therefore appear likely that individuals with alexithymia have an impaired ability to know the minds of others, which might contribute to difficulties in emotion regulation and interpersonal relations. However, the relation between alexithymia and mentalizing has only been sparsely investigated (for example, see Wastell and Taylor, 2002), and its neural basis remains to be examined.

We therefore investigated neuronal activation in individuals with high and low alexithymia during a ToM task using functional magnetic resonance imaging (fMRI). We hypothesized that alexithymia would be associated with decreased neuronal activity in the medial prefrontal cortex (MPFC), the temporo-parietal junction (TPJ), and the temporal pole (TP), a neuronal activity pattern characteristic of subjects who are known to have a deficit in the capacity for mentalizing, such as patients with Asperger’s syndrome (Castelli et al., 2002; Frith and Frith, 2003).

### Materials and method

#### Subjects

We screened 310 college students (105 males and 205 females) for alexithymia using a self-administrated questionnaire, TAS-20 (Taylor et al., 2003; Komaki et al., 2003). Individuals with high or low TAS-20 total scores ($n = 20$, score $> 60$ and $n = 18$, score $< 39$, respectively) were selected to obtain two groups that were maximally divergent on alexithymia. This yielded 38 volunteers (30 females, 8 males; 19 to 22 years of age, mean age = 20.4 years, SD = 0.938). All subjects gave written informed consent. The study was approved by the local ethics committees and conducted in accordance with the Declaration of Helsinki.

All subjects were interviewed by two medical doctors (psychiatry and psychosomatic medicine) using The Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998; Otsubo et al., 2005). All subjects were confirmed to have no major medical or neurological or psychiatric disorders and no history of psychiatric and psychosomatic disorders, including depression and anxiety disorders. Subjects were also administered the National Adult Reading Test (NART; Nelson and O’Connell, 1978; the Japanese version was developed by Matsuoka et al., 2002) to assess an intelligence quotient (IQ), which has been reported to correlate with alexithymia. All subjects were right-handed, confirmed by the Edinburgh Handedness Inventory (Oldfield, 1971).

All 38 subjects completed the psychological assessment (described above), the structured interview for alexithymia (SIBIQ, see below; Arimura et al., 2002; Sriram et al., 1988) and the fMRI study. The TAS-20 and SIBIQ scores are shown in Table 1. The 38 subjects were included in the fMRI study for correlational analyses between regional brain activities in response to mentalizing stimuli and the scores on the psychological measurements.

Some of the subjects had TAS-20 and SIBIQ scores that were discrepant on the two measures of alexithymia. To maximize the likelihood that subjects were correctly classified with regard to alexithymia, four participants with high TAS-20 and low SIBIQ scores and 4 with low TAS-20 and high SIBIQ scores were discarded. This yielded an Alexithymia (ALEX) group ($n = 16$, 3 males, mean(SD) age = 20.2 (1.0)) and a Non-alexithymia (NonALEX) group ($n = 14$, 2 males, mean (SD) age = 20.8 (0.89)), described in Table 1. These groups were used to examine group differences of the effects of the mentalizing stimuli on brain activity as measured by the fMRI and are to be distinguished from the sample of 38 subjects based on TAS-20 scores only.

#### Questionnaires

In the initial screening for high and low alexithymia, each subject had completed the 20-item Toronto Alexithymia Scale (TAS-20: Taylor et al., 2003; Komaki et al., 2003). The TAS-20 uses a five-point Likert response scale and has a three-factor structure consisting of ‘difficulty identifying feelings’, ‘difficulty describing feelings’, and ‘externally oriented thinking’.

To confirm each subject’s high or low level of alexithymia, we conducted a modified Structured Interview based on the Beth Israel Hospital Psychosomatic Questionnaire for alexithymia (SIBIQ; Arimura et al., 2002; Sriram et al., 1988). Originally, the SIBIQ was developed to ask about patients’ feelings related to their physical or psychiatric symptoms. For interviewing our non-patient

### Table 1

<table>
<thead>
<tr>
<th>Feature</th>
<th>min–max, mean (SD)</th>
<th>One sample ($n = 38$)</th>
<th>NonALEX ($n = 14$)</th>
<th>ALEX ($n = 16$)</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAS-20 Total</td>
<td>26–74, 51.2 (16.5)</td>
<td>26–38, 34.1 (3.7)</td>
<td>61–74, 66.1 (4.5)</td>
<td>−21.28</td>
<td></td>
</tr>
<tr>
<td>Difficulty in identifying feeling</td>
<td>7–32, 18.0 (8.1)</td>
<td>7–19, 10.6 (3.7)</td>
<td>19–32, 24.7 (3.9)</td>
<td>−10.33</td>
<td></td>
</tr>
<tr>
<td>Difficulty in describing feeling</td>
<td>5–25, 15.4 (6.2)</td>
<td>5–18, 9.6 (3.9)</td>
<td>15–24, 20.1 (2.4)</td>
<td>−9.16</td>
<td></td>
</tr>
<tr>
<td>Externally oriented thinking</td>
<td>9–30, 17.9 (5.1)</td>
<td>9–21, 13.9 (3.3)</td>
<td>13–30, 21.4 (4.0)</td>
<td>−5.57</td>
<td></td>
</tr>
<tr>
<td>SIBIQ Total</td>
<td>18–70, 42.2 (16.7)</td>
<td>18–56, 31.5 (11.8)</td>
<td>25–70, 52.2 (14.1)</td>
<td>−4.39</td>
<td></td>
</tr>
</tbody>
</table>

All $T$ values are significant $P < 0.001$; TAS-20: 20-item Toronto Alexithymia Scale, SIBIQ: the modified Structured Interview based on Beth Israel hospital psychosomatic Questionnaire for alexithymia, ALEX: Alexithymia group, NonALEX: Non-alexithymia group.
The remaining 120 volumes were used for the analyses. The sequence were discarded because of non-steady magnetization, and fMRI volume acquisitions. The first five volumes of each fMRI condition. Each epoch lasted 20 s, equivalent to five whole-brain epochs: a ToM task condition and a control random-movement condition matrix, 256; 1NEX field of view, 31.5 cm; slice thickness, 1.23 mm).

The Interpersonal Reactivity Index (IRI; Davis, 1996; Aketa, 1999) was used to evaluate empathic ability. The IRI evaluates four domains related to empathy and representation of other’s mental states: ‘Empathic Concern’ consisting of feeling emotional concern for others; ‘Perspective Taking’ which refers to cognitively taking the perspective of another; ‘Fantasy’, consisting of emotional identification with characters in books, films etc.; and ‘Personal Distress’ which refers to having negative feelings in response to the distress of others.

**Stimuli for fMRI measurements**

We conducted one fMRI session with each subject to assess the neural network associated with mentalizing, using eight silent visual animations for ToM (Abell et al., 2000; Castelli et al., 2000, 2002; Ohnishi et al., 2004; http://www.icn.ucl.ac.uk/dev_group/research.htm). We used two types of animations (4 ToM animations with two triangles acting like humans and 4 control animations with two triangles moving randomly), which were modified to be matched in length (20 s) without diminishing the meaning of the originals. The basic visual characteristics in terms of shape, overall speed, and orientation changes were as similar as possible in all the conditions.

**Scanning method and procedure**

All MRI measurements were performed with a 1.5 T Magnetom Vision plus MRI scanner (Siemens, Erlangen, Germany) using a standard head coil. Cerebral activation was measured with fMRI using blood–oxygen-level-dependent (BOLD) contrast. After automatic shimming, a time series of 125 volumes was obtained using single-shot gradient-refocused echo-planar imaging (TR 4000 ms; TE 60 ms; flip angle 90°; inter-scan interval 4 s; in-plane resolution 3.44 × 3.44 mm; FOV 22 cm; contiguous 4 mm slices to cover the entire brain). For structural imaging, a three-dimensional volumetric MRI was acquired using a T1-weighted gradient echo sequence that produced a gapless series of thin sagittal sections using an MPRage sequence (TE/TR, 4.4/11.4 ms; flip angle, 15°; acquisition matrix, 256 × 256; 1NEX field of view, 31.5 cm; slice thickness, 1.23 mm).

The fMRI protocol was a block design with two kinds of epochs: a ToM task condition and a control random-movement condition. Each epoch lasted 20 s, equivalent to five whole-brain fMRI volume acquisitions. The first five volumes of each fMRI sequence were discarded because of non-steady magnetization, and the remaining 120 volumes were used for the analyses. The subjects viewed the stimuli projected onto a screen, ~50 cm from the subject’s head, through a mirror attached to the head coil.

Before each fMRI experiment, subjects were told to watch the animations and think about what the triangles were doing and thinking. After each scan, the subjects again watched the same stimuli on a computer screen outside the scanner and were asked to recall what they had thought, during the fMRI scan, that the triangles had been doing. Then the subjects were asked to tell the experimenter what they had thought, in response to the experimenter’s neutral question: “What was happening in this animation?” Their verbal descriptions were coded on two different dimensions (based on the criteria of Castelli et al., 2000, 2002): ‘Intentionality’ (the degree of appreciation of mental states), and ‘Appropriateness’ (how well and correctly the underlying script was captured). The Intentionality score reflected the use of mental state terms, with scores ranging from 0 (non-deliberate action) to 5 (deliberate action aimed at affecting another’s mental state). The Appropriateness score measured the understanding of the event depicted in the animations, as intended by the designers (0 to 3).

Detailed scoring procedures are given in the report by Castelli et al. (2000).

**Data analyses**

We used Statistical Parametric Mapping (SPM2, Wellcome Department of Imaging Neuroscience, London) to realign EPI images and co-register them to the subjects’ T1-weighted MR images. Then the T1 images were transformed to the anatomical space of a template brain whose space is based on the MNI (Montreal Neurological Institute) stereotactic space. The parameters for the transformation were applied to the co-registered EPI images. The normalized images were smoothed using an 8-mm FWHM Gaussian kernel. A first fixed level of analysis was computed subject-wise using the general linear model with the hemodynamic response function modeled as a boxcar function whose length covered one animation of a particular type. To test the hypotheses about regionally specific effects of the mentalizing animation condition, the estimates were compared by means of linear contrasts for each epoch. The resulting set of voxel values for each contrast constituted a statistical parametric map of the t statistic SPM(t).

Anatomic localization is presented using the Talairach coordinate system (Talairach and Tournoux, 1988). First-level contrasts were then introduced in a second-level random-effect analysis (Friston et al., 1999) to allow for population inferences.

Main effects of mentalizing stimuli were computed using a conjunction analysis for the one-sample t tests for the alexithymia and the non-alexithymia group, which yielded SPM(t), subsequently transformed to the unit normal distribution SPM(μ). A voxel-level and a cluster-level spatial extent threshold of P < 0.05 corrected for multiple comparisons (False discovery rate, t > 2.76) were used to identify mentalizing-related brain areas.

To compare the alexithymic group (n = 16) and the non-alexithymic group (n = 14) within the regions activated in the present conjunction analysis, a two-sample t test was used. To avoid type-II errors, we conducted a hypothesis-driven approach for the regions that had been identified as important components associated with the mentalizing-related tasks (i.e., medial prefrontal cortex (MPFC), temporo-parietal junction in superior temporal sulcus (TPJ), and temporal pole and neighbor peri-
groups are shown in Table 3 and Fig. 1. We found no remarkable
to the control random task by conjunction analysis between the two
fMRI measurement
no group differences.
empathic concern, and scored significantly higher on IRI personal
to understand the meaning, we created regions of interest (ROI: 10 mm radius sphere) centered on each peak coordinate in the
the present one-sample t test and calculated individual mean contrast
values of each ROI using MarsBaR software (http://marsbar.
sourceforge.net/). Correlation coefficients between these individual
mean values and psychological variables were calculated.

Results

Behavioral measures

There was no difference between the groups in the ratio of male
to female subjects (13 alexithymia females to 3 alexithymia males
and 12 non-alexithymia females to 2 non-alexithymia males; χ² =
0.168, Fisher’s Exact probability = 1.00, two tailed), so scores for
men and women were combined (see Table 2). The alexithymia
group scored significantly lower than the non-alexithymia group
on ToM intentionality, appropriateness, IRI perspective taking and
empathic concern, and scored significantly higher on IRI personal
distress. Scores on the reading test (JART; reflecting IQ) revealed
no group differences.

fMRI measurement

The results of neural activations during the ToM task compared
to the control random task by conjunction analysis between the two
groups are shown in Table 3 and Fig. 1. We found no remarkable
change in the analyses of neural activation after analyzing the data
only for the female subjects (analyses not shown). Activations in
the right MPFC, bilateral TPJs, and the right TP/Amy were
observed. Additionally activations were observed in the following
areas: occipital visual cortices, cerebellum, middle temporal gyri,
inferior/middle frontal gyri, thalamus, parietal precuneus, and
supplementary motor cortices.

In the group comparisons, there were no significant group
differences using a conservative statistical threshold (family-wise
error P < 0.05). However, significantly decreased activations
during the ToM task in the alexithymic group were noted in the
right medial prefrontal cortex (MPFC; BA10 (x, y, z) mm = (12,
65, 19), Z = 3.69, k = 25) and the right superior temporal area
(BA22 (x, y, z) mm = (46, −35, 2), Z = 4.16, k = 44) using a more
lenient statistical threshold {Z < 3.09 (P < 0.001), k > 20}. On the
other hand, there were no statistically significant increased
activations during the ToM task in the alexithymic group even
using a more lenient statistical threshold (uncorrected P < 0.005).
We extracted 4 clusters and selected 4 peak coordinates in
individual clusters based on a priori hypotheses (i.e., MPFC,
TPJ, and TP) and from the peaks of the clusters generated in the
conjunction analysis {right MPFC (x, y, z) mm = (14, 65, 19); right
TPJ (53, −48, 13); left TPJ (−57, −50, 17); right TP (50, 9,
−24)}. The two regions found in the group comparison were tested
by means of the SVC (10 mm sphere) centered on those
cordinates from the conjunction analysis, and the only region
within a 10 mm radius centered on each coordinate that had been
specified by an a priori hypothesis was a decrease in the right
MPFC in the alexithymia group (height and extent, P = 0.010,
0.015 corrected with SVC, respectively; Fig. 2).

Correlations between BOLD activity during the ToM task in
each ROI and psychological measurements (Table 4) revealed
positive correlations between appropriateness of understanding
of ToM scripts and BOLD activity in the right MPFC, left TPJ, and
right TP. BOLD activity in the right MPFC was significantly
 correlated with ToM appropriateness and IRI perspective taking
and negatively correlated with SIBIQ score. The highest correla-
tion among these was with perspective taking (see Fig. 3). BOLD
activity in the right TP was positively correlated with mentalizing
and IRI personal distress scores.

Discussion

Our study demonstrates differences between individuals in
behavioral and neural responses to a mentalizing task as a function
of alexithymia. Alexithymia, a disturbance in self-awareness, was
associated with impairment in mentalizing and the related empathic
ability of perspective taking, the ability to see things from the point
of view of another person. Neural activity in the medial prefrontal
cortex was decreased in alexithymics, and activity in the same
region was closely related to perspective-taking scores. To our

Table 2

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NonALEX (n = 14)</td>
<td>ALEX (n = 16)</td>
</tr>
<tr>
<td>ToM scoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentionality</td>
<td>17.2 (1.9)</td>
<td>&gt;14.9 (3.5)</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>9.5 (1.7)</td>
<td>&gt;7.9 (1.9)</td>
</tr>
<tr>
<td>Interpersonal Reactivity Index (IRI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>19.5 (6.7)</td>
<td>18.0 (5.6)</td>
</tr>
<tr>
<td>Perspective taking</td>
<td>18.5 (4.7)</td>
<td>&gt;14.6 (3.5)</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>20.2 (3.7)</td>
<td>&gt;16.1 (5.0)</td>
</tr>
<tr>
<td>Personal distress</td>
<td>12.2 (3.8)</td>
<td>&lt;15.6 (4.1)</td>
</tr>
<tr>
<td>Expected IQ (JART)</td>
<td>109.5 (6.4)</td>
<td>111.7 (8.3)</td>
</tr>
</tbody>
</table>

NonALEX; non alexithymia group, ALEX; alexithymia group. Statistical significance; *P < 0.05. ns; not significant.
Table 3
The coordinates and T and Z scores for the brain areas activated with main effects of ToM animated stimuli; Conjunction analysis demonstrating overlap of the alexithymic and non-alexithymic group

<table>
<thead>
<tr>
<th>Brodmann area</th>
<th>Talairach coordinates x, y, z (mm)</th>
<th>Voxel P (FDR corrected)</th>
<th>T</th>
<th>Z</th>
<th>Cluster k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rt middle temporal gyrus</td>
<td>21</td>
<td>50 –31 0</td>
<td>0.000</td>
<td>10.79</td>
<td>6.73</td>
</tr>
<tr>
<td>Rt fusiform gyrus</td>
<td>37</td>
<td>42 –61 –9</td>
<td>0.000</td>
<td>10.42</td>
<td>6.61</td>
</tr>
<tr>
<td>Rt temporoparietal junction</td>
<td>22</td>
<td>53 –48 13</td>
<td>0.000</td>
<td>9.16</td>
<td>6.18</td>
</tr>
<tr>
<td>Rt middle temporal gyrus/temporal pole</td>
<td>38</td>
<td>50 9 –24</td>
<td>0.000</td>
<td>8.03</td>
<td>5.74</td>
</tr>
<tr>
<td>Lt parahippocampal gyrus/amygdala</td>
<td>*</td>
<td>24 –5 –17</td>
<td>0.014</td>
<td>3.43</td>
<td>3.11</td>
</tr>
<tr>
<td>Lt occipital lingual gyrus</td>
<td>18</td>
<td>–24 –91 –2</td>
<td>0.000</td>
<td>10.25</td>
<td>6.56</td>
</tr>
<tr>
<td>Lt cerebellum posterior pyramis</td>
<td>*</td>
<td>–14 –79 –30</td>
<td>0.000</td>
<td>8.05</td>
<td>5.75</td>
</tr>
<tr>
<td>Lt temporoparietal junction</td>
<td>22</td>
<td>–57 –52 16</td>
<td>0.000</td>
<td>5.50</td>
<td>4.51</td>
</tr>
<tr>
<td>Rt cerebellum posterior pyramis</td>
<td>*</td>
<td>16 –81 –31</td>
<td>0.000</td>
<td>6.70</td>
<td>5.13</td>
</tr>
<tr>
<td>Lt inferior frontal gyrus</td>
<td>45</td>
<td>–53 24 14</td>
<td>0.001</td>
<td>4.90</td>
<td>4.13</td>
</tr>
<tr>
<td>Rt thalamus</td>
<td>*</td>
<td>8 –11 4</td>
<td>0.001</td>
<td>4.61</td>
<td>3.95</td>
</tr>
<tr>
<td>Rt parietal precuneus</td>
<td>7</td>
<td>–6 –52 43</td>
<td>0.002</td>
<td>4.31</td>
<td>3.75</td>
</tr>
<tr>
<td>Lt parietal precuneus</td>
<td>7</td>
<td>–6 –50 41</td>
<td>0.011</td>
<td>3.55</td>
<td>3.19</td>
</tr>
<tr>
<td>Lt inferior frontal gyrus</td>
<td>9</td>
<td>–42 9 31</td>
<td>0.003</td>
<td>4.24</td>
<td>3.69</td>
</tr>
<tr>
<td>Rt middle frontal gyrus</td>
<td>6</td>
<td>–44 4 44</td>
<td>0.005</td>
<td>3.89</td>
<td>3.45</td>
</tr>
<tr>
<td>Rt medial prefrontal cortex</td>
<td>10</td>
<td>12 63 17</td>
<td>0.009</td>
<td>3.66</td>
<td>3.28</td>
</tr>
<tr>
<td>9</td>
<td>10 62 30</td>
<td>0.003</td>
<td>4.20</td>
<td>3.67</td>
<td></td>
</tr>
<tr>
<td>Rt supplementary motor area</td>
<td>8</td>
<td>6 49 44</td>
<td>0.005</td>
<td>3.96</td>
<td>3.5</td>
</tr>
<tr>
<td>Lt middle temporal gyrus</td>
<td>21</td>
<td>–57 –6 –11</td>
<td>0.011</td>
<td>3.55</td>
<td>3.2</td>
</tr>
<tr>
<td>Lt postcentral gyrus</td>
<td>2</td>
<td>–49 –28 60</td>
<td>0.013</td>
<td>3.45</td>
<td>3.12</td>
</tr>
<tr>
<td>Rt brainstem midbrain</td>
<td>*</td>
<td>6 –23 –2</td>
<td>0.015</td>
<td>3.41</td>
<td>3.09</td>
</tr>
<tr>
<td>Rt superior frontal gyrus</td>
<td>6</td>
<td>10 13 62</td>
<td>0.016</td>
<td>3.35</td>
<td>3.05</td>
</tr>
<tr>
<td>Lt fusiform gyrus</td>
<td>20</td>
<td>–42 –21 –24</td>
<td>0.017</td>
<td>3.34</td>
<td>3.04</td>
</tr>
<tr>
<td>Rt postcentral gyrus</td>
<td>2</td>
<td>69 –20 30</td>
<td>0.019</td>
<td>3.27</td>
<td>2.98</td>
</tr>
<tr>
<td>Lt frontal lobe/rectal gyrus</td>
<td>11</td>
<td>–4 30 –22</td>
<td>0.020</td>
<td>3.25</td>
<td>2.97</td>
</tr>
<tr>
<td>Lt uncus</td>
<td>20</td>
<td>–34 –7 –30</td>
<td>0.020</td>
<td>3.25</td>
<td>2.97</td>
</tr>
<tr>
<td>Lt thalamus</td>
<td>*</td>
<td>–10 –15 4</td>
<td>0.048</td>
<td>2.78</td>
<td>2.59</td>
</tr>
</tbody>
</table>

Rt; right, Lt; left, FDR; False discovery rate, k; cluster extent.

knowledge, this is the first study to investigate and demonstrate the neural substrates of ToM in alexithymia.

Previous imaging studies concerning alexithymia (Berthoz et al., 2002a,b; Kano et al., 2003; Mantani et al., 2005) have classified subjects as alexithymic or non-alexithymic solely based on the TAS-20. Self-administered questionnaires ask respondents to describe or estimate how they think, imagine, and feel about their own emotional states. However, the concept of alexithymia includes a core difficulty in identifying and describing the emotional state of the self. This may result in some inaccuracy in their ability to evaluate and rate themselves on their ability to identify or describe their feelings appropriately. In addition to using a self-administered questionnaire (TAS-20) as a screening instrument, we used a structured interview (SIBIQ) to select the sample. Together with the large sample obtained in the present study, this strategy enabled us to overcome a limitation of the TAS-20 and confidently classify subjects as alexithymic or non-alexithymic.

Behavioral measures

Our alexithymic group showed poorer ToM ability than the non-alexithymic group, based on the scoring of their verbal descriptions of the ToM animations. This is contrary to the findings of a study that used the False Beliefs subset of the Picture Sequencing Task (Wastell and Taylor, 2002). This discrepancy may be partly explained by the more rigorous selection criteria for the alexithymic sample in the present study, as discussed above. Another explanation is that the picture sequencing task is so simple that it may not have been able to detect subtleties in the evaluation of the mental states of others such as teasing, humor, joking, and faux pas (Nakamura et al., 2000). The performance in the Wastell and Taylor (2002) study may have been subject to ceiling effects, since only implicit but not explicit mentalizing was impaired. The movie stimuli in the present study included complex mental states such as persuading, bluffing, mocking, and surprising the other. The results of our behavioral measures suggest that alexithymia, impaired self-awareness, is associated with decreased mentalizing ability.

Fitzgerald and Molyneux (2004) proposed an overlap between alexithymia and Asperger’s syndrome, which is an autistic spectrum disorder characterized by an impairment in mentalizing. Autistic spectrum disorders have been reported to be associated with alexithymia (e.g., Hill et al., 2004). Although alexithymia occurs in individuals who do not have autistic spectrum disorders, our findings suggest that even in these alexithymic individuals a deficit in mentalizing may be present. It is possible that the nature of the mentalizing deficit may differ in the two groups, with the deficit in the autistic spectrum disorders being more pervasive or severe.

Furthermore, we found that perspective-taking ability and empathic concern were decreased in the alexithymic group, while the tendency for experiencing personal distress was increased. Guttman and Laporte (2002) also reported that alexithymic individuals had higher levels of IRI personal distress and lower levels of perspective taking and fantasy. The factors of perspective taking and empathic concern reflect mature empathy, whereas personal distress may reflect a more immature
tendency to identify with the other (Guttman and Laporte, 2000, 2002). The personal distress scale is associated with social dysfunction, fearfulness, uncertainty, emotional vulnerability, shyness, and anxiety. High personal distress was characterized by concern about evaluation by others and with lowered concern for others (Davis, 1983). Hence, alexithymics with high personal distress scores might not have mature empathetic ability but rather could easily be affected by others’ mental state. This notion is consistent with the fact that, unlike the other subscales of the IRI, personal distress has been shown to correlate positively with measures of antisocial behavior and aggression (Beven et al., 2004; Davis, 1996), and that antisocial personality disorder is associated with alexithymia (high TAS score; Sayar et al., 2001). Note that a potential weakness of the present study is the absence of information regarding depressive and anxiety states, which have been reported to correlate with alexithymia (Honkalampi et al., 2001; Kojima et al., 2003; Muller et al., 2003; Saarijarvi et al., 2001). However, we confirmed by using the structured interview that none of the subjects were diagnosed with either depression or an anxiety disorder.

Brain activations

In the fMRI study, we replicated the loci (TPJ, TP adjacent to amygdala, and MPFC) that have been reported in previous studies of mentalizing (Castelli et al., 2000, 2002; Frith and Frith, 2003). TPJ activity reflects representation of goal-directed action states and is involved in the processing of detecting agency through biological-motion (Frith and Frith, 2003). Samson et al. (2004) showed that TPJ sustains not only low-level social perception but also higher-level social
inference of someone else’s belief, such as false-belief reasoning. Saxe and Kanwisher (2003) showed that the role of the TPJ in understanding other people is specific to reasoning about the content of mental states, and thinking about beliefs is also subserved by the TPJ (Saxe, 2005).

These functions in the TPJ are supposed to support ToM abilities. The TP is concerned with generating the semantic and emotional context for the material currently being processed (Frith and Frith, 2003). The amygdala can process social information of various kinds, not just judgments of others’ emotional states, and may underlie or partially overlap with ToM functions (Stone et al., 2003).

MPFC activity in the context of a mentalizing task has been studied by many researchers (e.g., Berthoz et al., 2002a,b; Brunet et al., 2000; Castelli et al., 2000, 2002; Ferstl and von Cramon, 2002; Fletcher et al., 1995; Goel et al., 1995; Gallagher et al., 2000; Macrae et al., 2004; Ochsner et al., 2004; Schultz et al., 2003; Vogeley et al., 2001). Frith and Frith (2003) proposed that the MPFC adjacent to the paracingulate region (anterior to posterior portion of rostral ACC) associated with mentalizing tasks is activated whenever people are attending to certain states of the self or others. This area is concerned with the representation of the mental states of the self and others decoupled from physical state representations. The correlation that we observed between perspective-taking ability and the degree of neural activity in the MPFC region in response to mentalizing tasks in the present study is consistent with the Frith’s proposal. Furthermore, Mitchell et al. (2005) showed that greater MPFC activation accompanied judgments specifically about targets’ psychological states, compared to body parts, regardless of whether the target was a person or not (e.g., a dog). That report is consistent with our results that only triangles-moving-like-humans evoked MPFC activation in the subjects. Decety and Jackson (2004) pointed out that studies that investigated other-perspective versus self-perspective consistently involved the areas including the frontopolar cortex and medial prefrontal cortex when the participants adopted the perspective of another person. These neuroimaging studies about the functions in the MPFC together suggest that the area of the MPFC activated in

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Correlation coefficients between activities related to ToM in each ROI and scores of psychological measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rt MPFC</td>
</tr>
<tr>
<td>ToM scoring</td>
<td></td>
</tr>
<tr>
<td>Intentionality</td>
<td>0.26</td>
</tr>
<tr>
<td>Appropriateness</td>
<td></td>
</tr>
<tr>
<td>Structured Interview for BQ (SIBIQ)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>−0.36*</td>
</tr>
<tr>
<td>Interpersonal Reactivity Index (IRI)</td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>0.25</td>
</tr>
<tr>
<td>Perspective taking</td>
<td><strong>0.38</strong></td>
</tr>
<tr>
<td>Empathic concern</td>
<td>0.25</td>
</tr>
<tr>
<td>Personal distress</td>
<td>0.27</td>
</tr>
<tr>
<td>Expected IQ (JART)</td>
<td>−0.08</td>
</tr>
<tr>
<td>Rt; right, Lt; left, MPFC; medial prefrontal cortex, TPJ; temporoparietal junction, TP; temporal pole.</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>P &lt; 0.05.</td>
</tr>
<tr>
<td>**</td>
<td>P &lt; 0.01.</td>
</tr>
<tr>
<td>***</td>
<td>P &lt; 0.005.</td>
</tr>
</tbody>
</table>

Fig. 2. Brain images of the comparisons showing less regional cerebral activation in individuals with alexithymia (n = 16) in response to the ToM task than those without alexithymia (n = 14). The sagittal (a) and axial (b) orthogonal views of the brain are shown for the cluster with less activation in the alexithymic group compared to the non-alexithymic group in the right medial prefrontal cortex (MPFC) (BA10, Talairach coordinates (x, y, z) mm = (12, 65, 19), peak Z = 3.69, cluster k = 25). The bars on the right show the range of t scores for statistical parametric mapping.

Fig. 3. The correlation between IRI perspective-taking scores and the activities in the right medial prefrontal cortex (MPFC) in one-sample (n = 38). The figure illustrates a significant positive correlation between the perspective-taking scores of the Interpersonal Reactivity Index (IRI) and right MPFC activity in response to the ToM task vs. control stimuli in one-sample (n = 38). The correlation coefficient (r) at the voxel (x = 14, y = 65, z = 21) = 0.48, P = 0.0012, R² = 0.229.
mentalizing tasks is indicative of the representation of the mental states of both the self and others decoupled from outward reality and from each other (i.e., meta-representation). This meta-representation process is a key to the higher cognitive ability of comprehending the minds of others; e.g., adopting the perspective of others as separable and distinct from one's own perspective.

We found relative inactivity in the right MPFC in alexithymic individuals. Neuroimaging studies in which alexithymics processed emotion-laden stimuli (Berthoz et al., 2002a,b; Kano et al., 2003; Lane et al., 1998), including another study of Japanese subjects (Kano et al., 2003), demonstrated hypofunction in the medial frontal area in neighboring dorsal anterior cingulate cortex.

Our finding that alexithymia is associated with compromised comprehension of others’ mental states suggests that comprehension of the self and the other are closely related. Indeed, Ochsner et al. (2004) have demonstrated that evaluation of the emotional states of self and others relies on a network of common mechanisms centered on the MPFC. Jackson et al. (2005) reported that medial prefrontal activity was related to taking both self and others’ perspective when viewing hands and feet in painful situations. Furthermore, individuals with autistic spectrum disorders have showed less activation in the MPFC in response to the same mentalizing task as that used in the present study (Castelli et al., 2002), and the MPFC has been reported to be associated with the impaired mentalizing in autism (Goel et al., 1995; Happé et al., 1996; Juhasz et al., 2001). In our study, individuals with high alexithymia, compared to those with low alexithymia, showed less activation in the MPFC in response to the mentalizing task. Thus, there could be a common neural component for both normal individuals with high alexithymia and autistic people with impaired mentalizing ability, which also indicates a common component of understanding the self and others.

To identify and describe one’s own feelings as distinct from those of another person, one needs a third-person perspective different from the self per se which is full of emotion (i.e., enabling the self to be viewed as an object). To comprehend the mind of another, one should adopt the perspective of the other person. Therefore, the ability to generate perspectives that are different from the self (i.e., decoupling or meta-representation) is a crucial common component for understanding the mental states of self and others. A deficit in this mechanism should result in both alexithymia and impaired mentalizing and should be associated with a tendency to take on the other’s distress continguously and personally as one’s own. This may be why alexithymia is associated with deficits in empathic ability (Gutman and Laporte, 2000, 2002; Rastam et al., 1997) and in reading emotion from posed facial expressions (Lane et al., 1996; Mann et al., 1994; Pandey and Mandal, 1997; Parker et al., 1993, 2005). Furthermore, some psychiatric disorders associated with alexithymia (e.g., autistic spectrum (Berthoz and Hill, 2005; Hill et al., 2004; Frith, 2004), schizophrenia (Cedro et al., 2001; Maggini and Raballo, 2004a,b; Stanghellini and Ricca, 1995; Todarello et al., 2005; van ’t Wout et al., 2004), borderline personality (Gutman and Laporte, 2002), and psychopathy (Haviland et al., 2004)) have one common component: an ambiguous boundary between the self and others, including proneness to take on the emotional states of others with poor decoupling. That is why individuals with these disorders cannot take a third-person perspective outside the self, which is critical for regulating emotions. Indeed, Fonagy (2000) has asserted that a deficit in the capacity for reflection is the fundamental deficit in borderline personality disorder.

From a developmental viewpoint, Frith and Happé (1999) noted that there is little evidence from the developmental literature to suggest that mental states are attributed to the self before they are attributed to others (e.g., children do not pass the self’s belief question in the Smartsies test before passing the other’s belief question). They emphasized a common representational mechanism for attributing mental states to self and others. Leslie (1987) also proposed that meta-representation is necessary for attribution of any mental state, including a false belief, and is necessary equally for self and other attribution. In infancy, before clear mental representation of the self emerges, the perception of emotional signals arising externally (i.e., the earliest ToM task) and the subjective experience of internal emotion are undifferentiated. Gradually the child comes to know his or her own emotional experience by virtue of feedback from the mother or other caregivers (Lane, 2000; Stern, 1985; Taylor et al., 1997). On the other hand, Frith and Frith (1999); Frith (1996) proposed that preexisting abilities that are relevant to mentalizing include the ability to distinguish between actions of the self and actions of others, so that skills that contribute to mentalizing ability also advance self-awareness. Nevertheless, the development of representations of the self and others are inseparable and parallel to each other. The results of the present study raise the possibility that some individuals experience a fundamental developmental deficit in this domain resulting in an impairment in knowing and communicating how one feels, which is labeled alexithymia.

The temporal pole (TP) has been reported to code for mnemonic and contextual information (Nakamura and Kubota, 1995), discrimination of familiar faces and scenes (Nakamura et al., 2000), personal and famous names (Sugiura et al., 2006), conditions evoking sadness (e.g., Beauregard et al., 2001; Eugene et al., 2003; Gillath et al., 2005; Levesque et al., 2003), and the retrieval of emotional memories (Dolan et al., 2000; Shin et al., 1999). Our study showed that TP activity was associated with mentalizing scores, indicating that the TP is associated with some aspect of the mentalizing function, as a semantic and mnemonic process (not perspective taking), given the previous studies and the Friths’ hypothesis (2003). On the other hand, TP activity was correlated with personal distress scores as a potentially immature empathizing tendency (taking on another’s emotional state without decoupling). The meta-representation process that produces the self’s and the other’s perspectives separately might not be included in the function of the TP area, but instead included in the MPFC. Given the reports that have shown the role of the TP for processing and referencing mnemonic information familiar to oneself, the present association between TP activities, mentalizing, and personal distress scores suggests that the TP serves as an important neural component of ToM function by utilizing personal experiences of one’s own to comprehend the state of mind of others. If the ability to differentiate between the emotions of self and other is impaired, as in alexithymia, then personal distress in conjunction with impaired perspective taking may lead to inaccuracies in the attribution of mental states in the other.

Conclusion

Our findings demonstrate that alexithymia, a deficit in the ability to identify and describe the feeling states of the self, is related to impaired mentalizing (comprehending the mind of others), which in turn is associated with hypoactivity in the
MPFC. The deficit in mentalizing that we observed is associated with impairment in the higher cognitive ability to take a perspective different from the self, a skill that may be essential for the comprehension of the mental states of both self and others. The results also suggest that there is a common component, such as perspective taking, involved in both self-awareness and mentalizing. Impairment of this component may help to explain the impairment in emotional regulation characteristic of alexithymia. Additional research is needed to examine the extent to which these findings contribute to a more complete understanding of those psychiatric disorders that have been linked to alexithymia, including autistic spectrum disorders, schizophrenia, and borderline and psychopathic personality disorders.

Acknowledgments

This study was supported by the Research Grant (17A-3) for Nervous and Mental Disorders from the Ministry of Health, Labour and Welfare, Japan.

We thank Dr. Uta Frith for her generous donation of the animation to the project.

References


