‘Theory of mind’ impairments and their relationship to executive functioning following frontal lobe excisions

Andrea D. Rowe,1 Peter R. Bullock,3 Charles E. Polkey3 and Robin G. Morris2

1Department of Psychology and 2Neuropsychology Unit, Institute of Psychiatry and 3Academic Neurosurgery, King’s College Neuroscience Centre, London, UK

Correspondence to: Andrea Rowe, Department of Psychology, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK
E-mail: a.rowe@iop.kcl.ac.uk

Summary
It has been suggested that mental states play an important role in determining behaviour and that mental state attributions (‘theory of mind’) underlie the ability to understand and predict other peoples’ behaviour. Theory of mind was investigated in 31 patients with unilateral frontal lobe lesions (15 right-sided and 16 left-sided) by comparing their performance with that of 31 matched control subjects. The ability to infer first- and second-order beliefs was tested by requiring subjects to listen to stories in which a protagonist acted upon a false belief. Both patient groups exhibited significantly impaired performance on the two theory of mind measures. Both frontal lobe groups also exhibited a range of deficits in tests of executive functions, but analyses revealed that these seemed to be independent of theory of mind impairments. These findings are discussed in terms of the hypothesis of a specialized, adaptive brain system underlying theory of mind reasoning ability, and are related to observed difficulties in social functioning among patients with frontal lobe damage.

Keywords: theory of mind; executive functions; frontal lobe excisions

Abbreviations: ACoA = anterior communicating artery; ANOVA = analysis of variance; AS = Asperger’s syndrome; LF = left-sided frontal lobe excisions; LSD = least squares difference; RF = right-sided frontal lobe excisions; RHD = right hemisphere damage; SAH = subarachnoid haemorrhage; ToM = theory of mind; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WCST = Wisconsin Card Sorting Test; WMS = Wechsler Memory Scale

Introduction
Impairment in social cognition has long been recognized as a commonly observed effect of frontal lobe damage. In particular, case studies such as those of patients A. (Brickner, 1936), K.M. (Hebb, 1945), J.P. (Ackerly and Benton, 1948), E.V.R. (Eslinger and Damasio, 1985), D.T. (Eslinger et al., 1992) and G.K. and M.H. (Price et al., 1990) have provided rich and striking descriptions of a seemingly heterogeneous array of behavioural changes associated with social difficulties. These include insensitivity to social cues, egocentrism, indifference to the opinions of others, lack of restraint, diminished foresight, impaired self-monitoring, a tendency to exhibit inappropriate affect, and social withdrawal.

The underlying cognitive mechanisms and neuroanatomical correlates of social behaviour are largely unknown and so such deficits have remained difficult to categorize and quantify. Many patients with frontal lobe damage exhibit impaired social behaviour demonstrate, but only minor impairment on standard neuropsychological tests, including those held to be sensitive to frontal lobe dysfunction. An alternative approach to this problem may be found in investigations of impaired social cognition in individuals with autism. In this field of research, the concept of a ‘theory of mind’ (ToM) has been developed (Baron-Cohen et al., 1985; Leslie, 1987). ToM refers to the ability to represent mental states, such as beliefs and intentions, and is distinct from the ability to represent ‘real’ states of affairs. It enables us to make attributions and to reason about mental states and, in doing so, to understand and predict the behaviour of other people (Premack and Woodruff, 1978).

Although this general topic has a long history (Baldwin, 1897; Piaget, 1955; Premack and Woodruff, 1978), interest in ToM is largely due to the assertion that an impairment in social cognition, caused by a specific deficit in an
independent, innate module underlying the development of ToM, is responsible for the core deficits in autism (Baron-Cohen et al., 1985; Leslie, 1987). Using the classic ‘false belief’ paradigm devised by Wimmer and Perner to investigate ToM development in normal children (Wimmer and Perner, 1983), Baron-Cohen and colleagues found that autistic children failed to predict accurately the action of a protagonist directly after being provided with information that implied strongly that the protagonist held a mistaken belief (Baron-Cohen et al., 1985). Subsequent studies have provided empirical support for impaired ToM in autism. However, the notion of a specific, independent cognitive deficit underlying these problems has been challenged. Executive dysfunction in autism has been reported (Steel et al., 1984; Rumsey, 1985; Rumsey and Hamburger, 1988, 1990), and there has been debate about the relationship between executive dysfunction, ToM deficits and their primacy in autism and the relationship between frontal lobe damage and autism (Ozonoff et al., 1991a, b; Hughes and Russell, 1993).

It is possible that ToM and executive functions are independent and are associated in autism only because of proximity of the respective underlying neuroanatomical systems. However, other possibilities exist. ToM may be a primary deficit, with failure in the development of ToM impeding the development of executive control. Alternatively, executive dysfunction may result in a secondary deficit in the ability to exert control over one’s own mental states. Hughes and Russell favour the latter alternative (Hughes and Russell, 1993). They acknowledge the likelihood of a real impairment in the ability to ‘mentalise’ in autism, but they argue that ToM tests are simply types of executive functioning tests, i.e. impaired performance reflects deficits in general problem-solving skills, such as cognitive flexibility and response inhibition. The difficulty with this proposal is that individuals with Asperger’s syndrome (AS), a mild subtype in the spectrum of autistic disorders, exhibit executive dysfunction and yet are able to solve ToM problems (Ozonoff et al., 1991b). However, the extent to which the finding of Ozonoff and colleagues can be regarded as conclusive is unclear because it contrasts with their own observation that people with AS fail to display any use of ToM ability in informal conversation.

The study reported here investigated the relationship between ToM ability and executive functioning by examining the performance of a large group of patients with focal lesions of the frontal lobes on a false belief test and a range of tests of executive functions. An association between some executive functions and social competence, assessed using a self-report inventory, has been reported in such groups (Grattan et al., 1994). Furthermore, studies using functional neuroimaging to investigate ToM ability indicate that the prefrontal cortex is critically involved (Baron-Cohen et al., 1994; Fletcher et al., 1995; Goel et al., 1995). However, in the first group study of ToM in 10 subjects with frontal lobe damage, Stone and colleagues reported high performance on a simple first-order and second-order ToM test and impairment only on a ‘faux pas’ test in a subgroup of five subjects with more extensive lesions incorporating bilateral orbitofrontal cortex (Stone et al., 1998).

In the present study, the group of patients with frontal lobe excisions was sufficiently large and varied, in terms of site and size of lesions, for it to be feasible to look for effects of size, region and laterality of the lesions on ToM test performance. In addition to the tests of executive functions, the experimental design incorporated a number of other measures with which to constrain interpretations of ToM test performance. Measures internal to the false belief test included the standard control questions to check comprehension and memory for story details. As patients with right hemisphere damage (RHD) are thought to fail ToM tests because of a type of pragmatic impairment other than a difficulty in conceptualizing another person’s mental state (Siegal et al., 1996), questions designed to elicit inferences other than mental state inferences were also included. This provided an opportunity to examine the relationship between performance on these questions and the ToM test questions, which are essentially questions requiring references about mental states. Thus, the study was designed to determine whether first- and second-order ToM is impaired after prefrontal lesions and to examine the nature of any such impairments, in particular to determine their relationship to impairments on standard tests of executive functions.

Methods

Subjects

The subjects for this study included 31 adult neurosurgical patients with unilateral frontal lobe lesions who had undergone surgery at King’s Neurosciences Centre, London (formerly the Maudsley Hospital Neurosurgical Unit, London). Only patients with lesions exclusive to the frontal lobes were selected, and these lesions were identified and verified using the neurosurgeon’s drawings and notes made at the time of operation, and neuroimages produced by pre- and postoperative CT scanning and MRI. The neurosurgeon’s drawings made after examination of these documents, which depict the site and extent of the lesions, are shown in Figs 1 and 2. Patients gave informed consent and the study was approved by The Institute of Psychiatry Bethlam and Maudsley Trust Ethical Committee.

The patients were tested a minimum of 3 months postoperatively (mean 77 months, SD 60 months). Fifteen patients (six males and nine females) had right frontal lobe lesions and 16 (eight males and eight females) had left frontal lesions. In the right frontal lobe (RF) group, eight patients had a resection or a lobectomy incorporating an eliptogenic focus to treat pharmacologically intractable epilepsy, five had a tumour removed (four had a meningioma excised and one had an ogliodendroglioma excised), and two had a right
Fig. 1 Diagrams illustrating the extent of excisions (in black) in the RF group.
Fig. 2 Diagrams illustrating the extent of excisions (in black) in the LF group.
Table 1 Patient characteristics

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Aetiology</th>
<th>Lesion location</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF1</td>
<td>M</td>
<td>50</td>
<td>Meningioma</td>
<td>Dorsolateral +</td>
</tr>
<tr>
<td>RF2</td>
<td>M</td>
<td>65</td>
<td>Meningioma</td>
<td>Medial +</td>
</tr>
<tr>
<td>RF3</td>
<td>F</td>
<td>26</td>
<td>Epilepsy</td>
<td>Orbital +</td>
</tr>
<tr>
<td>RF4</td>
<td>M</td>
<td>28</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF5</td>
<td>F</td>
<td>58</td>
<td>Meningioma</td>
<td></td>
</tr>
<tr>
<td>RF6</td>
<td>F</td>
<td>70</td>
<td>ACoA aneurysm and SAH</td>
<td>Dorsolateral +</td>
</tr>
<tr>
<td>RF7</td>
<td>M</td>
<td>29</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF8</td>
<td>M</td>
<td>37</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF9</td>
<td>F</td>
<td>25</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF10</td>
<td>F</td>
<td>42</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF11</td>
<td>F</td>
<td>23</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF12</td>
<td>M</td>
<td>29</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>RF13</td>
<td>F</td>
<td>52</td>
<td>Meningioma</td>
<td></td>
</tr>
<tr>
<td>RF14</td>
<td>F</td>
<td>47</td>
<td>ACoA aneurysm and SAH</td>
<td></td>
</tr>
<tr>
<td>RF15</td>
<td>F</td>
<td>22</td>
<td>Oligodendroglioma</td>
<td>Medial +</td>
</tr>
<tr>
<td>LF1</td>
<td>F</td>
<td>33</td>
<td>Epilepsy</td>
<td>Dorsolateral +</td>
</tr>
<tr>
<td>LF2</td>
<td>M</td>
<td>58</td>
<td>Oligodendroglioma</td>
<td>Medial +</td>
</tr>
<tr>
<td>LF3</td>
<td>F</td>
<td>45</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>LF4</td>
<td>M</td>
<td>28</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>LF5</td>
<td>M</td>
<td>27</td>
<td>Haematoma: focal HI</td>
<td></td>
</tr>
<tr>
<td>LF6</td>
<td>M</td>
<td>37</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>LF7</td>
<td>M</td>
<td>48</td>
<td>Oligodendroglioma</td>
<td>Medial +</td>
</tr>
<tr>
<td>LF8</td>
<td>F</td>
<td>24</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>LF9</td>
<td>F</td>
<td>74</td>
<td>Meningioma</td>
<td></td>
</tr>
<tr>
<td>LF10</td>
<td>F</td>
<td>53</td>
<td>ACoA aneurysm and SAH</td>
<td>Medial +</td>
</tr>
<tr>
<td>LF11</td>
<td>M</td>
<td>36</td>
<td>Haematoma: focal HI</td>
<td>Dorsolateral +</td>
</tr>
<tr>
<td>LF12</td>
<td>M</td>
<td>33</td>
<td>Meningioma</td>
<td></td>
</tr>
<tr>
<td>LF13</td>
<td>F</td>
<td>64</td>
<td>Meningioma</td>
<td></td>
</tr>
<tr>
<td>LF14</td>
<td>F</td>
<td>59</td>
<td>ACoA aneurysm and SAH</td>
<td></td>
</tr>
<tr>
<td>LF15</td>
<td>F</td>
<td>42</td>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>LF16</td>
<td>M</td>
<td>27</td>
<td>Cerebral abscess</td>
<td></td>
</tr>
</tbody>
</table>

ACoA = anterior communicating artery; SAH = subarachnoid haemorrhage; HI = head injury.

The performance of the frontal experimental groups was compared with that of 31 healthy control subjects who had reported having no history of psychiatric or neurological disease. They were recruited to match the frontal experimental groups for sex, age, education and premorbid Full Scale IQ, as estimated by the National Adult Reading Test—Revised (Nelson and Willison, 1991) (Table 2). One-way analysis of variance (ANOVA) confirmed that the three groups were well matched in terms of current (pro-rated) Verbal IQ [F(2, 59) = 0.28, P = 0.76], years of education [F(2, 59) = 1.65, P = 0.20] and estimated IQ [F(2, 58) = 0.44, P = 0.64].

A battery of standardized neuropsychological tests was administered to all subjects (Tables 3 and 7). These were chosen to measure intellectual function, memory and executive functioning, both to provide background information on the patients and to explore the extent to which putative ToM deficits could be attributed to basic forms of cognitive impairment. The three groups did not differ significantly in terms of current (pro-rated) Verbal...
Table 2 Subject demographics

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex (M:F)</th>
<th>Age (years)</th>
<th>Education</th>
<th>NART FSIQ</th>
<th>Months since surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF (n = 15)</td>
<td>6:9</td>
<td>40.2 (15.9)</td>
<td>12.0 (2.2)</td>
<td>102.4 (12.7)</td>
<td>96.4 (58.8)</td>
</tr>
<tr>
<td>LF (n = 16)</td>
<td>8:8</td>
<td>44.2 (14.9)</td>
<td>13.0 (3.1)</td>
<td>104.5 (14.0)</td>
<td>60.5 (59.0)</td>
</tr>
<tr>
<td>NC (n = 31)</td>
<td>13:18</td>
<td>42.9 (15.4)</td>
<td>11.5 (2.3)</td>
<td>105.7 (8.3)</td>
<td>_</td>
</tr>
</tbody>
</table>

Data are mean (standard deviation). F = female; M = male; NART FSIQ = National Adult Reading Test—Revised, Full Scale IQ; NC = normal controls.

Table 3 Performance on Wechsler Adult Intelligence Scale—Revised

<table>
<thead>
<tr>
<th></th>
<th>RF</th>
<th>LF</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal IQ (pro-rated)</td>
<td>89.73 (12.44)</td>
<td>93.40 (13.75)</td>
<td>93.45 (6.73)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>8.1 (3.0)</td>
<td>9.5 (2.7)</td>
<td>8.7 (2.3)</td>
</tr>
<tr>
<td>Comprehension</td>
<td>8.1 (2.9)</td>
<td>9.7 (3.4)</td>
<td>8.7 (2.1)</td>
</tr>
<tr>
<td>Similarities</td>
<td>9.0 (2.4)</td>
<td>8.1 (3.1)</td>
<td>8.9 (2.3)</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.9 (2.0)*</td>
<td>8.4 (2.1)*</td>
<td>9.7 (1.9)</td>
</tr>
</tbody>
</table>

Data are age-scaled mean score (standard deviation). *Performance significantly worse than normal controls (NC) in post hoc comparisons using LSD (P < 0.05).

IQ, as estimated from performance on the Vocabulary, Comprehension, Similarities and Digit Span subtests of the Wechsler Adult Intelligence Scale—Revised (WAIS-R) (Wechsler, 1981), [F(2,59) = 0.73, P = 0.49] (Table 3).

Materials and procedure

Theory of mind test

To test ToM, stories were constructed about characters involved in routine activities, with implications regarding the protagonist’s false belief, and were followed by a series of questions. They were designed to test the ability to solve problems involving either (i) first-order attributions of false belief (of the form ‘A thinks X’), or (ii) second-order attributions of false belief (of the form ‘A thinks B thinks X’).

A total of 12 stories were constructed, six for the first-order false belief test and six for the second-order false belief test. The topics for the stories were finding a parking space, going to a restaurant, borrowing library books, paying a bill, doing housework, having the car serviced, being discharged from hospital, getting up in the morning, doing some redecorating, arranging a surprise party, going on holiday and going grocery shopping. Each story was followed by a ToM test question and three control questions.

(i) False belief test question. This was designed to elicit a response that demonstrated the ability to make inferences about another individual’s mental state, namely, that a protagonist held a false belief.

(ii) Inference question. The aim of this question was to assess the ability to draw inferences that did not involve reasoning about another individual’s mental state.

(iii) Fact question. This was posed to determine whether subjects had understood relevant events that, in actuality, stood in contrast to others and, in doing so, implied the false belief of the protagonist.

(iv) Memory question. This was used to assess whether memory for story details was approximately intact.

Examples of both types of story test, involving first- and second-order attributions of false belief, are provided in Appendix I. The subjects received the following instructions for the false belief test: ‘I am going to read you a number of different stories. I would like you to listen carefully because when I am finished I am going to ask you some questions. In order to answer these questions you will need to think about what has happened in the story and then draw your own conclusions.’

The stories and questions were then presented orally and responses were recorded.

Executive functioning

In order to investigate the relationship between putative ToM impairment and executive functions, a set of tests of executive functioning was administered, covering four main areas: (i) cognitive initiation or selection was tested with the Controlled Oral Word Association Test (Spreen and Benton, 1977), using F, A and S as the cue letters; (ii) response inhibition/interference was assessed with the Stroop Test (Stroop, 1935; Trenerry et al., 1989); (iii) mental flexibility was assessed with both the Trail-Making Test (Reitan and Wolfson, 1993) and the Wisconsin Card Sorting Test (WCST) (Heaton, 1981); and (iv) monitoring and organization were assessed with a simple experimental test—a verbal, externally ordered task adapted from Petrides and colleagues (Petrides et al., 1993).

Memory

To determine whether verbal memory impairment was a causative factor in terms of ToM impairment, the Wechsler Memory Scale (Wechsler, 1945) Logical Memory test of
A. D. Rowe et al.

Fig. 3 Performance on first-order theory of mind questions. The black column = LF; the stippled column = normal controls.

prose recall was administered. Working memory/attentional impairment was screened for with the Digit Span subtest of the WAIS-R.

Integration/sequencing
In order to control for possible differences in the ability to integrate and sequence material, the Picture Arrangement subtest of the WAIS-R was also administered.

Design
The background neuropsychological tests and the experimental test were administered in the same fixed order. However, within the experimental test, the presentation of first- and second-order false belief stories was alternated and blocked into four sets, each consisting of three stories. Each set was, therefore, made up of either two first-order false belief stories and one second-order false belief story or two second-order false belief stories and one first-order false belief story. In order to counterbalance the presentation of the different story sets, a 4 × 4 Latin square was used and repeated across each subject group. In a further attempt to offset any learning effects, the background tests were slotted between set presentations, following a fixed format, such that no set immediately preceded another.

Results
First-order theory of mind

Theory of mind
In an initial analysis, the responses to the ToM questions were scored as correct if the subject made an appropriate and correct first-order belief attribution (in the form ‘A thinks X’ or ‘A does not know Y’), or incorrect, if no appropriate belief attribution was made. A correct response was scored 1 and an incorrect response 0; the scores were averaged across stories and transformed into an overall percentage correct score (Fig. 3). A one-way ANOVA comparing the three groups showed a highly significant difference \[ F(2,59) = 21.15, P < 0.0001 \]. Post hoc comparisons between the individual group pairings were made using the least squares difference (LSD) method. This showed that the RF group’s performance did not differ significantly from that of the LF group, but that overall percentage correct scores for both patient groups were significantly lower than that of the control group.

Each incorrect response was categorized as one of four types of error.

(i) In ‘Incorrect Belief Attribution’, the subject made a belief attribution, but in a mistaken fashion so that it provided an inadequate explanation. For example, an attribution about the wrong character in the story was made, or the facts about the attribution were incorrect. (ii) In ‘Correct Mental State Attribution’, the subject failed to explain the behaviour of the protagonist in terms of their mistaken belief, but merely inferred another mental state appropriate to the scenario. (iii) ‘Incorrect Mental State Attribution’ was similar to (ii), but the mental state inference was about the wrong character in the story or was not appropriate to the scenario. In other words, it was irrelevant to the story. (iv) An ‘Inappropriate Response’ was, for example, a response that reiterated the facts of the story without attempting to answer the question appropriately.

In order to establish the reliability of response scoring and categorization, all responses for 28 of the subjects were rated independently by two people, one of whom had no knowledge of the group to which the subjects belonged. Overall, there was 98% inter-rater agreement.

The proportions of errors for the different categories were calculated, and are shown in Table 4. In the RF group, the Inappropriate Response category accounted for the largest proportion of errors, and the Correct Mental State Attribution category for the next lowest proportion. A very similar pattern was seen for the LF group. For the control group, the number of errors was so small overall that it made the pattern less reliable, but the Inappropriate Response and Correct Mental State Attribution categories also accounted for the highest proportions of errors.

Inference question. The response to each inference question was defined as being correct (1 point) or otherwise (0 points). The percentage correct scores for each group are given in Fig. 4. ANOVA showed a significant difference between the groups overall \[ F(2,59) = 4.44, P < 0.05 \]. Post hoc analyses with LSD tests showed that the percentage correct score for the LF group was significantly lower than that of the control group; this was the only significant difference between the individual groups.

Fact question. The responses were defined as correct or incorrect. The percentage correct scores are given in Fig. 4. ANOVA revealed no significant difference between the groups overall.
Table 4 Proportions for error response categories for first-order false belief task

<table>
<thead>
<tr>
<th>Error category</th>
<th>RF</th>
<th>LF</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect Belief Attribution</td>
<td>2.56 (1.11)</td>
<td>2.70 (1.04)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>Correct Mental State Attribution</td>
<td>23.08 (10.00)</td>
<td>16.22 (6.25)</td>
<td>33.37 (3.23)</td>
</tr>
<tr>
<td>Incorrect Mental State Attribution</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>5.58 (0.54)</td>
</tr>
<tr>
<td>Inappropriate Response</td>
<td>74.36 (32.33)</td>
<td>81.08 (31.25)</td>
<td>61.05 (5.91)</td>
</tr>
</tbody>
</table>

Proportion of all responses shown in parentheses. NC = normal controls.

Fig. 4 Performance on first-order false belief control questions. Cross hatched columns = RF; black columns = LF; stippled columns = normal controls.

Fig. 5 Performance on second-order theory of mind questions. The cross-hatched column = RF; the black column = LF; the stippled column = normal controls.

Memory question. The percentage correct scores are given in Fig. 4. The ANOVA showed no significant group differences.

Covariate analysis. In order to determine whether the ability to draw inferences from the stories, to understand facts or to remember material affected the ability to answer the ToM questions, the ToM data were subjected to analysis of covariance in which the measures relating to these variables were controlled. When the inference question measure was covaried, the overall difference between the groups remained highly significant \( F(2,58) = 19.35, P < 0.0001 \). This was also true for the fact \( F(2,58) = 18.98, P < 0.0001 \) and memory question measures \( F(2,58) = 21.12, P < 0.0001 \). Post hoc comparisons were also made between the individual group pairings, in which a stringent Bonferroni correction (Holmes, 1979) was applied. (For all post hoc comparisons a stringent Bonferroni correction was applied in which the statistical significance level of \( P = 0.05 \) was halved for every comparison, providing a \( P \) value of 0.006.) In each case, the ToM deficit in the performance of the two patient groups, relative to that of the control group, remained highly significant \( P < 0.0001 \).

Second-order theory of mind

Theory of mind

The responses to the ToM questions for the second-order stories, like those for the first-order stories, were scored as correct (in the form of e.g. ‘A thinks B doesn’t know X’ or ‘A thinks B thinks Y’) or incorrect. The overall percentage correct scores for the three groups are provided in Fig. 5. One-way ANOVA showed a highly significant difference between groups \( F(2,59) = 23.55, P < 0.0001 \). Post hoc analyses with LSD tests revealed that the overall scores for the two patient groups were not significantly different, but that the scores for the RF and LF groups were both significantly lower than that of the control group.

Eight error categories were used to classify the incorrect responses for second-order stories. (i) In ‘Incorrect Second-Order Belief Attribution’, a belief attribution taking the form of an embedded representation (of the form ‘A thinks B doesn’t know X’) was made, but was incorrect. This was either because the facts of the attribution were wrong or because the attribution was not addressed to the relevant story character. (ii) In ‘Correct First-Order Belief Attribution’, the subject made a belief attribution of the form ‘A thinks X’ or ‘A doesn’t know Y’. This can be regarded as superficially correct, but it fails to demonstrate an understanding of the rationale for the belief, even though this is strongly implied in the story. (iii) ‘Incorrect First-Order Belief Attribution’ corresponded to the first error category for responses to first-order ToM stories (as incorrect belief attribution). (iv) ‘Correct Second-Order Mental State Attribution’ took a form similar to (i) but referred to a mental state other than the implied belief (e.g. a desire) of one of the characters in the story as an explanation (e.g. ‘A thinks B wants X’). (v) ‘Incorrect Second-Order Mental State Attribution’ took the same form as (iv) but was incorrect because the attribution was not to the relevant character, or the details of the
Table 5 *Proportions for error response categories for second-order false belief task*

<table>
<thead>
<tr>
<th>Error category</th>
<th>RF</th>
<th>LF</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect Second-Order Belief Attribution</td>
<td>1.54 (1.11)</td>
<td>2.90 (2.08)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>Correct First-Order Belief Attribution</td>
<td>41.54 (30.00)</td>
<td>37.68 (27.08)</td>
<td>39.99 (12.90)</td>
</tr>
<tr>
<td>Incorrect First-Order Belief Attribution</td>
<td>12.30 (8.89)</td>
<td>15.94 (11.46)</td>
<td>21.67 (6.99)</td>
</tr>
<tr>
<td>Correct Second-Order Mental State Attribution</td>
<td>1.54 (1.11)</td>
<td>1.45 (1.04)</td>
<td>13.33 (4.30)</td>
</tr>
<tr>
<td>Incorrect Second-Order Mental State Attribution</td>
<td>1.54 (1.11)</td>
<td>2.90 (6.08)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>Correct First-Order Mental State Attribution</td>
<td>10.77 (7.78)</td>
<td>8.70 (6.25)</td>
<td>3.35 (1.08)</td>
</tr>
<tr>
<td>Incorrect First-Order Mental State Attribution</td>
<td>0.00 (0.00)</td>
<td>1.45 (1.04)</td>
<td>4.99 (1.61)</td>
</tr>
<tr>
<td>Inappropriate Response</td>
<td>30.77 (22.22)</td>
<td>28.98 (20.83)</td>
<td>16.67 (5.38)</td>
</tr>
</tbody>
</table>

Proportions as comparative shares of all responses shown in parentheses. NC = normal controls.

Table 6 *Comparison of performance levels on first-order and second-order ToM questions for frontal subjects (n = 31)*

<table>
<thead>
<tr>
<th>Performance</th>
<th>First-order ToM stories</th>
<th>Second-order ToM stories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>High</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>

For first-order ToM stories, mean number of correct first-order belief attributions = 3.45, therefore frontal subjects who scored ≤3 were classified as performing at a low level and those who scored ≥4 were classified as performing at a high level. For second-order ToM stories, mean number of correct belief attributions (first and second order-see text) = 3.23, therefore frontal subjects who scored ≤3 were classified as performing at a low level and those who scored ≥4 were classified as performing at a high level.

Groups were collapsed to form one group. The mean number of Correct First-Order Belief Attributions scored for the six first-order ToM stories was then calculated (3.45) and used to define performance levels as high (score ≥4) or low (score ≤3). The mean score for responses classified as Correct Second-Order and First-Order Belief Attributions for the six second-order ToM stories was also calculated (3.23), and this was used to define performance levels for false belief attributions on the second-order ToM stories as high (score ≥4) or low (score ≤3). The number of frontal subjects scoring within these categories (those defined as high-level performers and those defined as low-level performers) on the first- and second-order ToM stories are shown in Table 6. This reveals that there was a substantial degree of overlap; the patient’s performance levels for false belief attribution
on first-order ToM stories tended to be predictive of their performance level for false belief attributions on second-order ToM stories.

**Inference question.** The responses to the inference questions were scored as either correct (1 point) or incorrect (0 points). Figure 6 shows the percentage correct scores. Overall, there was a significant difference between the groups \( F(2,59) = 4.45, P < 0.05 \). Again, post hoc LSD tests revealed that the percentage correct score for the LF group was significantly lower than that of the control group, and that this was the only significant group difference.

**Fact question.** Figure 6 shows the percentage correct scores for the three groups. ANOVA showed no significant differences overall.

**Memory question.** The percentage correct scores are given in Fig. 6. ANOVA showed that there were no significant differences between the groups.

**Covariate analysis.** In new analyses, by covarying measures of ability to draw inferences and to understand and remember details from the second-order stories, the extent to which these factors could account for ability to answer ToM questions was examined. The overall difference between the groups remained highly significant when covarying the inference question \( F(2,58) = 17.90, P < 0.0001 \), the fact question \( F(2,58) = 22.66, P < 0.0001 \) and the memory question measure \( F(2,58) = 24.48, P < 0.0001 \). Post hoc analyses, in which Bonferroni correction was applied, revealed that the deficit in the performance of the RF and LF groups, compared with the control group, remained highly significant in each case \( P < 0.0001 \).

**Effects of practice**

In order to study the effects of practice, within each group the scores for the ToM questions were collapsed according to order of presentation, such that three values were obtained for both first-order and second-order ToM stories (for the first and second, the third and fourth, and the fifth and sixth stories of each type presented). Using the Friedman test, no significant practice effects were found in any of the three groups. However, a trend on the second-order ToM stories for only the RF group \( \chi^2 (2) = 5.00, P = 0.08 \) was examined further. Wilcoxon tests revealed that, although there were no significant differences between the values that represented collapsed scores for the third and fourth and the fifth and sixth second-order ToM stories presented, the value for the first and second stories presented was significantly lower than the values for scores on the third and fourth \( Z = -1.99, P < 0.05 \) and the fifth and sixth stories presented \( Z = -2.24, P < 0.05 \).

**Executive functioning**

The performance data for the three groups on the tests of executive functioning are shown in Table 7. Comparisons between the groups were made using a series of ANOVAs. For the Controlled Oral Word Association Test, there was a significant effect of group \( F(2,59) = 7.75, P < 0.005 \). The LSD method was used for post hoc comparisons. This revealed that the two frontal groups performed significantly less well than the control subjects. There was also a significant difference between groups on the Stroop test \( F(2,59) = 7.34, P < 0.005 \); the LSD tests showed that only the performance of the LF group was impaired relative to the control group, and that the LF group also performed significantly worse than the RF group. The data for the alphabetical sequencing component (Part A) of the Trail-Making Test revealed an overall group effect \( F(2,59) = 9.99, P < 0.0005 \); LSD tests showed that both patient groups performed significantly less well than the control group. For the set-shifting component of the Trail-Making Test (Part B), there was also a group effect \( F(2,59) = 3.41, P < 0.05 \), but only the LF group were impaired in comparison with the control group. For the WCST, there was a significant difference between groups in the Number of Categories achieved \( F(2,59) = 3.19, P < 0.05 \), but only the LF group performing significantly less well than the control group. There was a trend for the Percentage of Perseverative Errors score \( F(2,59) = 2.50, P = 0.09 \), and the LSD analysis again indicated a performance deficit in only the LF group relative to the control group. The Failure to Maintain Mental Set score did not differentiate the groups. For the Externally Ordered task, there was a significant effect of group \( F(2,59) = 7.30, P < 0.005 \); both frontal groups performed significantly worse than the control subjects.

**Memory**

The average score for the Wechsler Memory Scale Logical Memory Test was computed. There was no significant group effect \( F(2,59) = 0.90 \), indicating an absence of impairment.
in the patient groups. For Digit Span, there was a significant difference between the groups \(F(2,59) = 5.15, P < 0.01\); LSD tests showed that the performance of both patient groups was significantly worse than that of the control group (Table 7).

**Integration/sequencing**

Mean scores on the WAIS-R Picture Arrangement subtest for the three groups were calculated and are shown in Table 7. ANOVA showed that there were no significant differences between the groups.

**Re-analysis of ToM with executive functioning and memory as covariates**

To determine whether the impairment in ToM could be related to the executive functioning or memory impairments, the ToM data were re-analysed using the executive functioning and memory measures as covariates.

**First-order theory of mind**

For all the analyses with executive functioning measures, the main effect of group for the ToM data remained highly significant \((P < 0.0001)\). Similarly, comparisons between individual groups using the stringent Bonferroni correction continued to show highly significant differences between patients and control subjects. Nevertheless, performance on the Stroop test contributed significantly to the variance in the main analysis of first-order ToM \(F(1,58) = 7.03, P < 0.05\). This was also true for the set-shifting component (Part B) of the Trail-Making test \(F(1,58) = 8.27, P < 0.01\) and the WCST Percentage of Perseverative Errors score \(F(1,58) = 7.74, P < 0.01\). The post hoc comparisons between the individual group pairings revealed that the main source of this variance was the performance of the LF group relative to the control group. On these, and additionally on the WCST Number of Categories achieved measure, a significant contribution to the variance was found only in the comparisons between the LF and control groups \([\text{Stroop}, F(1,45) = 8.50, P = 0.006; \text{Trail-Making Test Part B}, F(1,45) = 10.39, P < 0.006; \text{WCST Percentage of Perseverative Errors score}, F(1,45) = 19.02, P < 0.006; \text{WCST Number of Categories achieved}, F(1,45) = 8.87, P < 0.006]\).

For the memory measures, the ToM differences remained robust with Logical Memory as a covariate \(F(2,58) = 15.85, P < 0.0001\); post hoc comparison between groups showed the same results. For Digit Span, this was also true for the main effect \(F(2,58) = 14.46, P < 0.0001\) and post hoc analyses. In both cases, there was a significant contribution to the variance \([\text{Logical Memory}, F(1,58) = 4.22, P < 0.05; \text{Digit Span}, F(1,589) = 4.233, P < 0.05]\). However, the performance of neither the RF nor the LF group, in comparison with the controls, made significant individual contributions to this variance.

**Second-order theory of mind**

The main effect of group for the second-order ToM data also remained highly significant \((P < 0.0001)\) when re-analysed with the executive functioning measures as covariates. This was also the case in post hoc comparisons between individual groups with the Bonferroni correction. In the main analyses, the WCST Percentage of Perseverative Errors scores \(F(1,58) = 9.04, P < 0.005\) and the WCST Number of Categories achieved \(F(1,58) = 6.30, P < 0.05\) contributed significantly to the variance. Among the individual group pairings, only the Percentage of Perseverative Errors score in the comparison between the LF and control group \(F(1,45) = 10.57, P < 0.006\) made a significant contribution to the variance.

Similarly, for analyses with the memory measures, the main effect of group for the second-order ToM data remained \([\text{Logical Memory}, F(2,58) = 20.34, P < 0.0001; \text{Digit Span}, F(2,58) = 17.75, P < 0.0001]\). No significant contributions

| Table 7 Executive functioning, memory and integration/sequencing test scores |
|-----------------|-----------------|-----------------|
| RF              | LF              | NC              |
| Controlled oral word association test | 29.33 (12.57)* | 27.80 (9.55)* | 39.55 (10.73) |
| Stroop          | 94.14 (20.01)   | 73.73 (34.26)* | 101.39 (16.84) |
| Trail-Making Test Part A | 43.07 (16.17)* | 41.93 (11.47)* | 29.23 (8.75) |
| Trail-Making Test Part B | 104.79 (38.51) | 125.52 (122.67)* | 72.19 (32.29) |
| WCST            |
| Number of Categories achieved | 4.00 (2.24) | 3.33 (2.47)* | 4.94 (1.81) |
| Percentage of Perseverative Errors | 18.96 (13.22) | 24.81 (21.43)* | 14.45 (11.13) |
| Failure to Maintain Set | 1.64 (1.12) | 1.20 (1.57) | 0.90 (1.08) |
| Externally ordered (% correct) | 60.07 (26.54)* | 49.88 (27.19)* | 75.77 (18.12) |
| Logical memory (WMS) | 21.12 (6.99) | 21.00 (7.61) | 23.32 (5.55) |
| WAIS-R          |
| Digit Span      | 7.93 (2.02)* | 8.38 (2.13)* | 9.74 (1.88) |
| Picture Arrangement | 8.80 (2.91) | 9.06 (2.54) | 10.19 (2.80) |

*Performance significantly worse than normal controls (NC) in post hoc comparisons using LSD (P < 0.05). Standard deviations shown in parentheses.
to the variance were found in the main analyses of second-order ToM or in post hoc comparisons between individual group pairings.

**Lesion effects**

Supplementary analyses were conducted in order to investigate the effects of site and extent of lesions within the patient groups on performance on the first- and second-order ToM test questions.

The $t$-test was used to compare overall percentage correct scores for both first-order and second-order stories for each class within the patient groups. Scores for all classes of patients with lesions encroaching on individual specific unilateral sectors were compared with those of patients with lesions in other sectors lateralized to the same hemisphere or otherwise, and (more generally) were also compared with those of all other patients. For example, percentage correct scores for patients with lesions incorporating the left dorsolateral sector ($n = 7$) were compared with those of patients in the LF group with lesions that did not encroach on the left dorsolateral sector ($n = 9$), and also with those of all patients (i.e. LF and RF) with lesions that did not encroach on the left dorsolateral frontal sector ($n = 24$) (Table 1). The classes, defined in terms of individual sectors, were also amalgamated so that all possible unilateral combinations of sectors were represented in a new set of classes. The $t$-test was then used to compare percentage correct scores for patients in each of these new classes with those of patients outside these classes. For example, scores for patients with the largest lesions, incorporating the left dorsolateral, medial and orbital sectors ($n = 7$) were compared with those of all others in the LF group, with lesions encompassing fewer sectors ($n = 9$), and with those of all other (i.e. LF and RF) patients ($n = 24$). No significant effects were observed.

In addition, effects of laterality and size of lesion were analysed separately. First, the two frontal groups were collapsed to form one group, and the mean scores were calculated for correct first-order belief attribution on the first-order ToM test questions and for correct second-order belief attributions on the second-order ToM test questions. The mean scores were then used to classify subjects as performing at a high or low level according to whether they scored above or below the relevant mean. With subjects reassigned to their original RF or LF group, Table 8 shows the number of RF and LF subjects scoring within each of these categories for the ToM test questions. The Chi-square test revealed that there was no significant effect of laterality on performance level on either the first or second-order ToM test questions. The two frontal groups were again collapsed and regrouped according to size of lesion. The mean diameter of the lesions was 3 cm; subjects with lesions of ≤3 cm were therefore categorized as having small lesions and those with lesions >3 cm in diameter as having large lesions. Within these categories, subjects were reclassified, according to their scores, as performing at a high or low level on the first- and second-order ToM test questions (Table 9). No significant effects of size were found with the Chi-square test.

**Discussion**

This study found that both the RF and the LF groups were significantly impaired in ToM ability, as assessed using first- and second-order false belief tests. The relation between performance on the ToM test and on control questions within the test, the measures of executive functioning, memory, integration and sequencing, and verbal IQ indicates that this ToM impairment was not artefactual and was independent.

The control questions in the false belief test and some of the other tests examined whether impaired ToM performance could be due to a deficit in memory. Frontal patients are impaired on certain tests of verbal memory, including digit span (Janowsky et al., 1989), free recall of unrelated words (Jetter et al., 1986; Janowsky et al., 1989) and temporal order memory (Milner, 1971; Shimamura et al., 1990). Nevertheless, such patients are unimpaired on story recall tests (Janowsky et al., 1989; Frisk and Milner, 1990), which are similar in type to the false belief stories. In the present study, frontal patients were also unimpaired on story recall. A deficit in digit span was found, but covariate analysis showed that this could not account for the ToM impairment. Unimpaired performance on the Picture Arrangement test further indicates that it cannot be explained in terms of poor integration and sequencing of story information. This is consistent with the lack of impairment in their responses to the memory and fact questions in the false belief test, which
were designed specifically to test comprehension of particular facts that contrast with the behaviour of the protagonist and are relevant to his or her implied false belief.

The non-mental state inference questions in the false belief test were included in order to address the suggestion that failure on ToM tests can be due to a type of pragmatic deficit other than impaired ability to conceptualize another person’s mental state. Siegal and colleagues reported that RHD subjects could predict behaviour on the basis of a protagonist’s false belief when questions were stated explicitly and unambiguously, but were impaired in their ability to answer ToM test and control questions, seeming to misunderstand, when questions were framed in less explicit language, such that their pragmatic purpose was only implied (Siegal et al., 1996). RHD adults are known to exhibit difficulties with figurative, non-literal language, and this has been explained in terms of impaired use of pragmatic cues in contextual information (i.e. social and situational information) that imply meaning. They are thought to have difficulties in integrating information and drawing inferences from implied propositions and, therefore, they fail to go beyond concrete, literal interpretations (Brownell et al., 1986; Foldi, 1987; Weylman et al., 1989). In interpreting their own findings, Siegal and colleagues concluded that RHD subjects failed ToM test questions and control questions when a literal interpretation of their meaning was insufficient because of such an impairment in pragmatic language comprehension.

A different patient population was the focus in the present investigation, but the inference questions were used to examine the ability to apprehend the meaning implied in contextual information in stories, and then to explore the relationship between this ability and the ability to answer ToM test questions. Here, the LF group and not the RF group exhibited impaired performance in answering inference questions appropriately for both first- and second-order ToM stories. However, covariate analyses revealed that this deficit was independent of, and therefore could not account for, impaired performance on ToM test questions. This suggests that these deficits reflect separate problems, at least in this patient group.

A range of tests of executive functions was employed to examine the relationship between subjects’ performance on these tests and the ToM test. Deficits in the frontal groups were found, which is consistent with findings in earlier studies. Both the LF and the RF groups demonstrated impaired cognitive initiation in verbal fluency (Perret, 1974; Miceli et al., 1981) and impaired monitoring on an externally ordered working memory test (Petrides et al., 1993). Both frontal groups were also impaired on Part A of the Trail-Making Test (simple sequencing), but only the LF group were impaired on Part B, which tests mental flexibility; this is consistent with other studies (e.g. Reitan and Tarsesh, 1959). Only the LF group were impaired on the Stroop test and the WCST (Categories Achieved and Percentage Perseverative Errors), which is also in accordance with previous findings (Milner, 1971; Drewes, 1974; Perret, 1974).

Associations between performance on tests of executive functions and performance on the ToM test might be predicted on the basis of other lines of evidence. Grattan and colleagues reported a correlation between performance on measures of mental flexibility and empathy in subjects with frontal lobe damage and in another group of patients with mixed lesions (Grattan and Eslinger, 1989; Grattan et al., 1994). Furthermore, Hughes and Russell argue that ToM tests are themselves tests of executive function: competition between the subject’s own knowledge of story details and the inferred false belief of the protagonist may elicit any difficulties the subject may have in inhibiting a prepotent response (Hughes and Russell, 1993). Deficits in executive functions, including impaired mental flexibility, are also reported in subjects with autism, a group who generally fail ToM tests and in whom impaired ToM is believed to explain the core social deficit (Ozonoff et al., 1991a; Hughes and Russell, 1993; Hughes et al., 1994). Indeed, in the present study we found correlations between performance on the ToM test and tests of executive functions, but only for the LF group (on the Stroop test, Part B of the Trail-Making Test, and the WCST for the first-order test, and only on the WCST Percentage of Perseverative Errors for the second-order test). However, the important finding from covariate analyses was that these deficits are not necessarily causally related. Deficit in executive functioning does not appear to account for the ToM impairment in either frontal group; instead, it seems to be an independent deficit. The deficits in executive functions in the LF group may be more importantly related to their impaired performance on the ToM inference questions. Executive functions are likely to be involved in the cognitive operations underpinning the ability to draw inferences from the pragmatic implications of a narrative, and studies have indicated that pragmatic impairment is associated with cognitive rigidity in RHD (Brownell et al., 1986) and autistic subjects (Ozonoff and Miller, 1996).

Further analyses were undertaken to examine the nature of the ToM impairment in the frontal groups. A practice effect was found only in the RF group, but this was just on the second-order ToM questions, for the first two stories presented relative to the subsequent four. The control subjects, from the outset, performed at a markedly higher level than both frontal groups, but practice may have enabled RF subjects to develop ‘non-ToM’ strategies in an attempt to solve the second-order problems. The use of non-ToM strategies has been suggested as an explanation for the discrepancy between the observed failure of AS subjects to exhibit ToM ability during informal conversation and their tendency to pass classic paradigm ToM tests (Ozonoff et al., 1991b). Such strategies do not seem to have been available to LF subjects as no such effect was found in this group.

With regard to the discrepancy in the behaviour of AS subjects, Bowler suggested that theirs might be a failure in the everyday application of ToM rather than competence (Bowler, 1992). This implies that ToM tests cue the use of an impaired ToM facility, guiding relevant cognitive processes that would otherwise be available automatically. If this were so, then it is conceivable that the second-order ToM stories
may act as a stronger trigger than first-order ToM stories. However, when the level of success for all frontal subjects on the first-order ToM test questions was related to their level of success on the second-order ToM test questions (defined in terms of correct first- and second-order belief attributions), evidence of improvement in subjects’ performance on the second-order ToM test questions was relatively limited. Instead, first-order ToM responses tended to predict performance on the second-order ToM test questions. Indeed, six of the 10 frontal subjects defined as performing at a low level on both the first- and the second-order ToM stories made no second-order belief attributions at all, while 10 of the 11 defined as performing at a high level on both the first- and second-order ToM stories made second-order belief attributions in >50% of their correct responses. This suggests that impaired ToM ability, as assessed in this patient population with this test, may largely reflect the degradation of a complex ToM system, with severity of impairment related to the degree of degradation.

A current topic of debate is whether ToM ability is critically underpinned by an innate, dedicated, domain-specific cognitive module or mechanism (e.g. The ToM mechanism defined by Leslie, 1987) with corresponding neurological specificity, or subserved solely by domain-general, distributed processes. Variability across subjects in performance on ToM tests has been reported previously (in an RHD group and in a normal control group; Winner et al., 1998), and this is consistent with the commonly observed variation among individuals in the level of functioning in independent processes (e.g. in spatial ability). In contrast, if content-independent processes are responsible, then this presupposes that proficiency is dependent upon experience as this is the only source from which to develop expertise and, in the case of ToM ability, this is, at least, not intuitively appealing. Evidence of a regular developmental sequence in understanding false beliefs in children (Wimmer and Perner, 1983; Perner and Wimmer, 1985), which seems to apply across cultures (Avis and Harris, 1991), satisfies some ontogenetic and phylogenetic criteria for establishing whether ToM ability is dependent upon a specialized, adaptive, cognitive system. Furthermore, the case of autism, which appears to be characterized by a specific ToM impairment (Baron-Cohen et al., 1985), with non-social reasoning intact (Baron-Cohen et al., 1993), does suggest a discrete module, or set of modules, vulnerable to selective malfunction.

The present study cannot, in isolation, illuminate whether the observed deficit may occur solely with prefrontal cortical lesions as opposed to damage to other brain areas. Nevertheless, functional imaging studies provide indirect support for the view that this area has a special role. These studies have reported a distinct pattern of activation associated with the performance of ToM tests, with critical areas in the prefrontal cortex, even though they are, collectively, inconclusive as to the specific prefrontal area or areas involved. Using PET with slightly different tasks, both Fletcher and colleagues and Goel and colleagues isolated activation specific to a ToM task condition in the left medial frontal gyrus (Fletcher et al., 1995; Goel et al., 1995). Fletcher and colleagues identified left medial area 8 and Goel and colleagues identified left medial area 9. This is at odds with a SPECT (single photon emission computed tomography) study by Baron-Cohen and colleagues that used a task with no inference component (Baron-Cohen et al., 1994). Here, increased activation during performance of a test of recognition of mental state terms was observed in the right orbitofrontal region (relative to the left orbitofrontal and right and left polar frontal regions). These differences in prefrontal areas of activation may simply reflect differences in task demands. However, the present study has not clarified this issue. Analyses that examined the effects of lateralization, size and site of lesions (in terms of areas of functional significance) failed to localize a specific prefrontal region involved in the attribution of beliefs. An effect of lateralization of lesions might have been expected, and has been reported in earlier studies of the performance of this patient population on other tasks (Miotto et al., 1996; Morris et al., 1997); however, the absence of effects of site or size of lesions is perhaps not surprising as many of the lesions were large and most encompassed quite a number of different cytoarchitectonically defined prefrontal areas. One possible explanation is that ToM ability itself may be too complex to be localized easily. C. Frith has speculated that ToM ability might be underpinned by a suite of frontal cortex modules, each adapted and specialized to represent a different ‘propositional attitude’ that takes the form of an ‘intentional stance’ (Dennett, 1987), i.e. mental states such as believing, intending, knowing, and being aware (of something) (C. Frith, 1996).

In practice, the consequences of a ToM impairment or, in particular, the inability to profit from knowledge which implies that another person holds a false belief, would include misunderstanding other people’s behaviour when it is based on false belief. More generally, the ability to make, and take into account, attributions regarding other people’s mental states is crucial in interacting in a socially appropriate and sensitive manner. Difficulties in this regard have been reported repeatedly in case studies of patients with frontal lobe damage. Their antisocial behaviour has been described in terms of an apparent unintentional insensitivity to social rules and to the feelings of others; having ‘little sense of fairness towards others . . .’ but tending to feel victimized (Price et al., 1990); and being easily frustrated (Ackerly and Benton, 1948). Although they may be gregarious, patients with frontal lobe lesions tend to be socially isolated to varying degrees and seem to lack the wherewithal to arrive at sound judgements and decisions in social situations (Damasio et al., 1991; Saver and Damasio, 1991).

ToM impairment in patients with frontal lobe lesions can therefore explain social incompetence within the realm of ‘interactive’ social behaviour, as it does in the case of people with autism and Asperger’s syndrome (U. Frith,
1996). Even so, Stone and colleagues concluded that, on the basis of their high performance on simple, classic paradigm first- and second-order false belief tests and their impairment only on a ‘faux pas’ test, patients with frontal lobe damage have a more subtle ToM deficit than people with autism and AS (Stone et al., 1998). A possible alternative interpretation is that adults with acquired frontal lobe lesions may be better equipped to work out non-ToM strategies in order to pass such tests than those with developmental disorders. Representations of various elements of social knowledge seem to be preserved (Saver and Damasio, 1991; Swain et al., 1998) and some patients with frontal lobe lesions may be able to benefit from utilizing this in some way to compensate for ToM dysfunction when faced with the simplest of ToM problems in test settings. However, there is a wide range of behavioural manifestations of frontal lobe dysfunction, and ToM impairment clearly cannot account for all of these, nor is it likely to be responsible for all reported difficulties in social cognition. For example, differences have been observed on scales such as the Hogan empathy measure (Grattan et al., 1989) and Cornelius and Caspi’s EPSI (Everyday Problem-Solving Inventory) (Dimitrov et al., 1996), which assess social cognition in more general terms without distinguishing between social behaviours that are dependent upon the ability to make mental state attributions and those that are not.

In contrast, ToM tests are designed with the aim of isolating those aspects of social cognition associated with two-way reciprocal interactions that rely crucially on ToM ability. The false belief tests in the present study have facilitated the demonstration of ToM impairment, in patients with lesions of the prefrontal cortex, and shown that this deficit is independent of non-mental state inferencing. Finally, within the context of this experimental design, the ToM deficit and executive functioning deficits in patients with frontal lobe lesions are not causally related. In conclusion, if these deficits are not fundamentally related, this adds support to the view that a specialized, discrete ToM module, or set of modules, is instantiated in the frontal lobes, and evidence that some of these deficits can co-occur, as in people with autism, must be due to the proximity of the respective significant underlying functional areas.

References
Baldwin JM. Social and ethical interpretations in mental development: a study in social psychology. New York: Macmillan; 1897.
Brickner RM. The intellectual functions of the frontal lobes: a study based upon observation of a man after partial bilateral frontal lobectomy. New York: Macmillan; 1936.


Appendix I

First-order theory of mind story

Chris and Ruth are doing their housework. Ruth is vacuuming the carpets upstairs and Chris has just finished cleaning the bathroom. He takes the linen basket downstairs to put the clothes in the washing machine. Ruth sees him and switches off the vacuum cleaner for a moment. Ruth says to Chris, ‘I’ve got some bed linen here to go into the wash, so if there’s any space left in the washing machine don’t switch it on yet. I’ll come downstairs when I’m finished.’

Chris puts the contents of the linen basket into the washing machine. There is still some space left so he leaves the door open. He pulls out the soap tray on the machine and opens the cupboard to find a small box of washing powder. Chris uses the scoop inside the box to transfer a measured amount of powder to the soap tray in the washing machine. He closes the soap tray and goes back upstairs to Ruth. She is standing at the top of the stairs and is about to start vacuuming the stair carpet. Chris says to Ruth, ‘Why don’t you take the bedding downstairs and get the washing on? I’ll vacuum the stairs.’

Ruth adds some of the bedclothes to the washing machine. She opens the cupboard and takes a measured scoop of washing powder and puts it into a ball-like container. Ruth puts this container inside the washing machine on top of the washing powder and closes the door. She switches the machine on.

When she comes back later, Ruth finds soap suds oozing from the washing machine.

Test question: Why is Ruth puzzled to find so many soap suds?

Examples of frontal subjects’ responses
(i) ‘Because they’ve both put powder in the machine.’
(ii) ‘Because there’s double the amount of soap . . . They both filled the machine with soap.’
(iii) ‘Because one lot’s been put in already.’

Examples of control subjects’ responses
(i) ‘She doesn’t know he’s already put soap in . . . She thinks he’s just loaded it.’
(ii) ‘She doesn’t know he’s already put soap in the top . . . She put it inside.’

Inference question: What chore must Ruth have been doing before she started vacuuming the bedrooms?

Example of subjects’ responses
(i) ‘Stripping and making the beds.’

Fact question: Did Chris put soap powder into the washing machine?

Memory question: What had Chris been doing upstairs before he put the washing on?

Second-order theory of mind story

Richard and Ann want to redecorate their spare bedroom. They have several wallpaper samples which they have brought home from the DIY shop. After choosing one, they decide to make a start the next weekend. Ann says she can collect the wallpaper on her way home from work next Friday evening.

On Friday morning Richard’s boss unexpectedly tells him he can take the afternoon off. When he gets home he phones Ann at work. Unfortunately she is not available. One of Ann’s colleagues tells him she is at a meeting and she is not sure when Ann will be back. Richard explains that he wants to tell her he has got the afternoon off and he is going to make a start on their redecorating. He says ‘Can you ask Ann to call me at home when she returns?’

By the time Ann gets back to her office it is after 5 p.m. and all her colleagues have gone home. As she collects her things she sees a note on her desk saying, firstly, that her husband called to tell her he had the afternoon off and was beginning their redecorating work and, secondly, that she should phone him at home. It is late and Ann wants to get to the DIY shop before 6 p.m. so she decides that, since she already has the message, she won’t bother to phone Richard and she’ll just rush down to buy the wallpaper.

At home Richard has been working very hard. He hasn’t heard from Ann and he wonders why her colleagues are so unreliable. He’s nearly finished stripping off the old wallpaper when he hears Ann coming through the front door. He rushes downstairs and says ‘You’re going to be surprised when I tell you what I’ve been doing this afternoon’.

Test question: Why does Richard say this?

Examples of frontal subjects’ responses
(i) ‘Because he had the afternoon off and he’d stripped the wallpaper off.’
(ii) ‘Because they were going to start the work together.’
(iii) ‘Because he’s been decorating, and she’s got the paper, and she comes home to find he’s been decorating.’

Examples of control subjects’ responses
(i) ‘Because he assumes she hadn’t got the message, and so he thinks she doesn’t know he’s been home and started.’
(ii) ‘He thinks she doesn’t know . . . He thinks she hasn’t got the message and doesn’t know he’s had the afternoon off.’

Inference question: Why do you think Ann needed to hurry to get to the DIY shop?

Example of subjects’ responses
(i) ‘Because it closed at six o’clock.’

Fact question: Did Richard’s message about having the afternoon off?

Memory question: Which room were they redecorating?