INVESTIGATING THE NEURAL CORRELATES OF

"THEORY OF MIND"

Helen Louise Gallagher
MSc, DCR(R)

Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, UK.

A thesis submitted to the University of London in partial fulfillment for the Degree of Doctor of Philosophy

Prepared under the supervision of Professor Chris D. Frith and Dr Cathy Price

June 2002
ABSTRACT

This thesis presents a series of studies that examine the neural correlates of “theory of mind”. This is the ability to attribute independent mental states to self and others in order to explain and predict behaviour. It is an automatic and universal function in humans and a fundamental element of social cognition. It has been proposed that theory of mind ability arises from an innate, dedicated, domain-specific, and possibly modular cognitive mechanism (Fodor, 1991; Leslie and Thaiss, 1992). This proposal gains particular support from studies of autism, a biologically based developmental disorder that appears to be characterised by a selective impairment in theory of mind. Frith (1991) has suggested that this impairment accounts for almost all of the abnormal social, communicative and imaginative behaviour manifest by individuals with autism.

This thesis describes five experiments investigating the behavioural and biological characteristics of theory of mind ability. These studies support Leslie’s notion of a dedicated cognitive mechanism, indicating that a circumscribed region of the anterior paracingulate cortex mediates this mechanism. The studies also demonstrate that this region forms part of a distributed neural network associated with social perception. The results from these studies are discussed in relation to the cognitive mechanisms underpinning our everyday ability to ‘mind-read’.
ACKNOWLEDGEMENTS

I would like to thank my supervisors, Chris Frith and Cathy Price, for their constant support, inspiration and enthusiasm during my PhD. Also, Uta Frith and Francesca Happé for their ideas and advice. I would also like to thank Richard Frackowiak for giving me this opportunity in the first place and his unwavering belief in me since. I’m grateful to my many colleagues without whom this thesis may never have been completed; Jenny Coull, Rebecca Elliott, Nick Ward, Sarah Blakemore, Cath Mummery, Geraint Rees, Eleanor Maguire, Paul Fletcher Rik Henson and John Ashburner also my fellow radiographers and Peter Aston, Ric Davies, Rachel Maddock, Chloe Hutton, Eric Featherstone and Chris Freemantle for their technical support.

I would especially like to thank my family, particularly my parents and sisters Jayne and Kirsty, for their endless love and support, and my friends Stephen, Jenny, Rebecca, Anna, Nick, Mary, Sarah, Sylke, Cath, Geraint and Paul for their kindness and encouragement.

Finally, this is for my beloved son, Matthew, who has been with me almost as long as this thesis and never fails to make me smile.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Contents</td>
<td>4</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction

1.1 Theory of mind                                                      | 17   |
1.2 The developing theory of mind                                       | 18   |
1.3 Theories of theory of mind                                          | 22   |
1.4 Theory of mind deficits                                             | 27   |
  1.4.1 Autism                                                          | 27   |
  1.4.2 Acquired deficits                                               | 28   |
1.5 Neuroimaging and theory of mind                                     | 29   |
  1.5.1 Neuroanatomical Imaging                                         | 29   |
  1.5.2 Functional Imaging                                              | 31   |
1.6 Functions of the anterior paracingulate cortex                       | 35   |
1.7 The amygdala and theory of mind                                     | 37   |
1.8 Questions explored in this thesis                                   | 38   |
1.9 Acknowledgement of assistance                                       | 40   |

## Chapter 2: Methods, analysis and experimental design in functional neuroimaging

2.1 Introduction                                                        | 41   |
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Functional neuroimaging methods</td>
<td>41</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The relationship between neuronal activity and cerebral blood flow</td>
<td>41</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Positron Emission Tomography (PET)</td>
<td>42</td>
</tr>
<tr>
<td>2.2.2.1</td>
<td>General overview</td>
<td>42</td>
</tr>
<tr>
<td>2.2.2.2</td>
<td>PET experimental design</td>
<td>43</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Functional Magnetic Resonance Imaging (fMRI)</td>
<td>45</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>General Overview</td>
<td>45</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>fMRI Experimental design</td>
<td>49</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Structural Magnetic Resonance Imaging</td>
<td>50</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Comparing PET and fMRI</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>Experimental designs</td>
<td>53</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Functional segregation and functional integration</td>
<td>53</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Subtraction designs</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Factorial designs</td>
<td>55</td>
</tr>
<tr>
<td>2.4</td>
<td>Statistical analysis</td>
<td>56</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Background</td>
<td>56</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Realignment</td>
<td>57</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Spatial normalisation and smoothing</td>
<td>58</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Statistical inference</td>
<td>59</td>
</tr>
<tr>
<td>2.5</td>
<td>Variation in sample size and gender distribution</td>
<td>61</td>
</tr>
</tbody>
</table>
Chapter 3: Reading the Mind in the Verbal and Nonverbal Domains

3.1 Introduction ................................................................. 63
3.2 Methods ........................................................................... 66
   3.2.1 Participants .............................................................. 66
   3.2.2 Tasks ........................................................................ 66
      3.2.2.1 Story comprehension task ................................. 67
   3.2.2.2 Cartoon Task ...................................................... 69
   3.2.3 Data Acquisition .................................................... 72
   3.2.4 Data Analysis .......................................................... 72
3.3 Results ............................................................................. 74
   3.3.1 Theory of Mind Stories versus Non-Theory of Mind Stories .................................................. 74
   3.3.2 Theory of Mind Stories versus Unlinked Sentences .......................... 74
   3.3.3 Non-Theory of Mind Stories versus Theory of Mind Stories .................................................. 74
   3.3.4 Non-Theory of Mind Stories versus Unlinked Sentences .................................................. 74
   3.3.5 Theory of Mind Cartoons versus Non-Theory of Mind Cartoons ........................................... 74
   3.3.6 Theory of Mind Cartoons versus Jumbled Pictures .................................................. 75
   3.3.7 Non-Theory of Mind Cartoons versus Theory of Mind Cartoons ........................................... 75
3.3.8 Non-Theory of Mind Cartoons versus Jumbled Pictures. 75
3.3.9 Conjunction: Theory of Mind Stories versus Non-Theory of Mind Stories and Theory of Mind Cartoons versus Non-Theory of Mind Cartoons. 75
3.3.10 Interaction: (Theory of Mind versus Non-Theory of Mind) x (story comprehension versus cartoons). 76
3.3.11 Interaction: (Theory of Mind versus Non-Theory of Mind) x (cartoons versus story comprehension). 77
3.4 Discussion 78
3.4.1 Medial prefrontal cortex (BA 8/9/32) 78
3.4.2 Temporo-parietal junctions (BA 39/40) 81
3.4.3 The right middle frontal gyrus (BA 6) 82
3.4.4 The medial parietal (precuneus) cortex (BA 7/31) 83
3.4.5 The left flocculus of the cerebellum and the fusiform 84
3.5 Conclusions 84

Chapter 4: Imaging the Intentional Stance
4.1 Introduction 86
4.2 Materials and Methods 90
4.2.1 Subjects 90
4.2.2 Data Acquisition 90
4.2.3 Data Analysis 91
4.2.4 Experimental Paradigm 92
4.2.5 Behavioural Data 95
## Contents

4.3 Results ................................................................. 96
   4.3.1 Verbal reports .............................................. 96
   4.3.2 Performance measures ................................. 99
   4.3.3 Brain activity .............................................. 99

4.4 Discussion ......................................................... 101
   4.4.1 Intentional stance ........................................ 101
   4.4.2 The role of anterior paracingulate cortex ....... 104

4.5 Conclusions ....................................................... 108

Chapter 5: Representation of Expressive and Instrumental Gestures in Normal and Autistic Children

5.1 Introduction ....................................................... 109

5.2 Part 1 ................................................................. 112
   5.2.1 Methods ....................................................... 112
      5.2.1.1 Subjects .................................................. 112
      5.2.1.2 Task ....................................................... 112
      5.2.1.3 Data analysis ........................................ 113
   5.2.2 Results ....................................................... 114
   5.2.3 Discussion ................................................... 115

5.3 Part 2 ................................................................. 118
   5.3.1 Methods ....................................................... 118
      5.3.1.1 Subjects .................................................. 118
      5.3.1.2 Tasks ....................................................... 119
      5.3.1.3 Data analysis ........................................ 119
Chapter 6: The Neural Correlates of Gesture Representation: Expressive and Instrumental

6.1 Introduction ......................................................... 129

6.2 Methods ................................................................. 133
   6.2.1 Subjects ........................................................... 133
   6.2.2 Tasks .............................................................. 133
   6.2.3 Data Acquisition ............................................. 134
   6.2.4 Data Analysis .................................................. 137

6.3 Results ................................................................. 139
   6.3.1 Expressive vs. instrumental (conjunction) ........ 139
   6.3.2 Instrumental vs. expressive (conjunction) ....... 139
   6.3.3 Recognise vs. observe (conjunction) .............. 140
   6.3.4 Observe vs. recognise (conjunction) ............... 141
   6.3.5 Interaction: (recognise expressive (RE) – observe expressive (OE)) x (observe instrumental (OI) - recognise instrumental (RI)) ............................................. 142
   6.3.6 Interaction: (recognise instrumental(RI) – observe instrumental (OI)) x (observe expressive (RE) - recognise expressive (RE)) ............................................. 143

6.4 Discussion ............................................................ 144
LIST OF FIGURES

1-1 The sabotage and deception task ............................... 19
1-2 The Sally-Ann task .................................................. 21
2-1 Factorial design ....................................................... 56
3-1 Examples of Stories: Theory of Mind, Non-Theory of Mind
and unlinked sentences .................................................. 68
3-2 Examples of Cartoons: Theory of Mind, Non-Theory of Mind
and jumbled pictures ...................................................... 70
3-3 A statistical parametric map (SPM{Z}) as a maximum
intensity projection showing the areas where there was greater
activation to theory of mind stories and cartoons compared to
non-theory of mind stories and cartoons ............................ 77
3-4 Area of activation in the medial frontal cortex of a single
volunteer elicited by theory of mind stories and cartoons .... 80
4-1 Diagrammatic representation of the display and timing of the
stimuli ................................................................. 92
4-2 An SPM showing activation in the most anterior portion of the
paracingulate cortex elicited in the contrast of mentalising
versus rule solving in sagittal, coronal and axial views ......... 101
Contents

5-1 Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of the 4-year-old, 5-year-old and 6-year-old groups of normal children 115

5-2 Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of each group ................................................................. 121

5-3 Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of the 4, 5 and 6-year-old groups and the two autistic groups (verbal IQ > 6 and verbal IQ < 6) .................................................. 122

6-1 A statistical parametric map (SPM(Z)) as a maximum intensity projection showing the areas where there was greater activation to expressive gestures compared to instrumental gestures (conjunction) .............................................. 139

6-2 A statistical parametric map (SPM(Z)) as a maximum intensity projection showing the areas where there was greater activation to instrumental gestures compared to expressive gestures (conjunction) .............................................. 140
6-3 A statistical parametric map (SPM{Z}) as a maximum intensity projection showing the areas where there was greater activation in recognising compared with observing only (conjunction of expressive and instrumental gestures) ........ 141

6-4 A statistical parametric map (SPM{Z}) as a maximum intensity projection showing the areas where there was greater activation in observing only compared with recognising (conjunction of expressive and instrumental gestures) ........ 141

6-5 A statistical parametric map (SPM{Z}) as a maximum intensity projection showing the areas where there was greater activation in the interaction: (RE – OE) x (OI - RI). ............. 143

6-6 Sagittal and coronal views of activation in the right superior temporal sulcus elicited by recognition of expressive gestures (interaction: (RE – OE) x (OI - RI) ......................... 147

6-7 Sagittal and coronal views of activation in the anterior paracingulate cortex elicited by recognition of expressive gestures (interaction: (RE – OE) x (OI - RI) ......................... 150

6-8 Sagittal and coronal views of activation in the left inferior frontal gyrus and middle frontal gyrus elicited by recognition of instrumental gestures compared to expressive gestures (conjunction) ......................... 156
Contents

7-1 Points of maximum activity in the anterior paracingulate cortex associated with mentalising from the functional imaging studies described in this thesis .................. 166

7-2 Points of maximum activity in the anterior paracingulate cortex associated with mentalising from previous published studies of mentalising, mapped onto figure 7-1b ........... 167
# List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Regions of increased brain activity associated with theory of mind compared with non-theory of mind for i. story comprehension, ii. cartoons, and iii. story comprehension and cartoons</td>
<td>76</td>
</tr>
<tr>
<td>3-2</td>
<td>Regions of increased brain activity associated with theory of mind compared with non-theory of mind specific to i. story comprehension and ii. cartoons</td>
<td>77</td>
</tr>
<tr>
<td>4-1</td>
<td>Higher cognitive processes thought to be engaged in each condition</td>
<td>94</td>
</tr>
<tr>
<td>4-2</td>
<td>Regions of increased brain activity associated with mentalising compared with (i) rule solving and (ii) random generation</td>
<td>100</td>
</tr>
<tr>
<td>5-1</td>
<td>List of the 23 expressive and 23 instrumental gestures</td>
<td>113</td>
</tr>
<tr>
<td>5-2</td>
<td>Representation of instrumental and expressive gestures in normal children</td>
<td>115</td>
</tr>
<tr>
<td>5-3</td>
<td>Representation of instrumental and expressive gestures in children with autism</td>
<td>120</td>
</tr>
<tr>
<td>6-1</td>
<td>List of the 22 expressive and 22 instrumental gestures</td>
<td>136</td>
</tr>
<tr>
<td>6-2</td>
<td>Main effects comparison of expressive and instrumental gestures</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>Main effects comparison of observe and recognise</td>
<td>142</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>Interactions</td>
<td>143</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

1.1 Theory of mind.

One aspect of human behaviour that sets us apart from other primates is our superior social cognitive abilities. The most exceptional of these must be our ability to explain or predict the behaviour of others, almost as if we can read their minds. This ability is often referred to as having a "theory of mind" or mentalising. Our ability to mentalise is an automatic, high level and almost without exception, human function (Heyes, 1998). In these circumstances we can have no data from animal studies directly relevant to the brain systems underlying mentalising ability. The development of this ability is believed to be determined by an innate cognitive mechanism (Leslie, 1987) possibly dedicated and domain specific (Leslie and Thaiss, 1992; Fodor, 1992). Evidence from studies of autism support this theory. Autism is a biologically based disorder which appears to be characterised by a selective impairment in theory of mind (Frith et al. 1991). Interest in the brain basis of normal theory of mind, is fired by the hope of better understanding the neural systems which are abnormal in people with autism.

The term "theory of mind" was originally introduced by Premack and Woodruff (1978). It denotes the theories we as humans make about the mental states of others and can be defined as our ability to explain and predict the behaviour of others by attributing to them thoughts, knowledge, beliefs and desires independent
from our own. This ambiguous name has resulted in a number of other terms, including mentalising (Frith et al, 1991) or mind-reading (Happe, 1994) being used to refer to the same ability. Such terms will be used synonymously in this text.

A good illustration of what is meant by mentalising ability is a study by Sodian and Frith (1992). They investigated autistic, mentally retarded, and normal children's ability to deceive or obstruct an opponent from obtaining a sweet in a box (Fig 1-1). Each child was instructed to "always help the friend but never help the thief". This required either physically sabotaging the thief's attempts by locking the box or mentally deceiving the thief by pretending that the box was locked. When required to tell a lie, by saying that the box was locked when it wasn't, autistic children performed significantly worse than their controls, even taking into account mental age. However, they readily prevented a competitor from gaining a reward by physical manipulation i.e. locking the box and hiding the key. Their success on sabotage demonstrated that their failure on deception was not due to an inability to understand the task but to a failure to "understand the mind" of their opponent.

1.2 The developing theory of mind

As mentioned before, Autism is a biologically based disorder. It is characterised by a triad of impairments – problems in socialisation, communication and
imagination (Wing and Gould, 1979). It has been proposed that the co-occurrence of these three handicaps can be accounted for in an impairment of “theory of mind” ability or “mind-blindness”. The theory stems from the observation that autistic children do not spontaneously engage in pretend play. Pretence is good evidence that a child has not only primary representations of things as they really are in the world, but also metarepresentations which are separate from reality. If

Figure 1-1. The sabotage and deception task (Sodian & Frith 1992). In this task the child is shown two puppets; one is the friend and the other, the thief. A sweet is placed in a box and the child is told to “always help the friend, never help the thief”. In the sabotage task (a.) the child is told when the puppet is coming and asked what he will do. The child should either lock the box to prevent the thief from obtaining the sweet or leave it open to help the friend. In the deception task (b.) the box is unlocked and the child is asked what he will say when the puppet asks if the box is open. The child should lie to the thief and tell the truth to the friend.
autistic children do not show pretence because they cannot form metarepresentations, then they should also be incapable of understanding other mental states. In order to test this hypothesis an understanding of the normal development of theory of mind is necessary, about which there is now a good deal of evidence.

A normally developing child shows signs of understanding the minds of others as early as eighteen months when he engages in joint attention, protodeclarative pointing (Baron-Cohen, 1989) and pretend play (Leslie, 1987) and by the age of two he will understand the mental state of desire (Wellman and Woolley, 1990). Between three and four years the child starts to recognise that other people may hold beliefs and knowledge about the world that are different from his own (Wimmer and Perner, 1983). This can be demonstrated with standard false belief tests, such as the “Smartie test” (Perner et al. 1989) or the “Sally-Anne task” (Wimmer and Perner, 1983), which require the child to recognise that one character in the test has a mistaken belief about the actual state of affairs (See Fig 1-2). Normal children from age 4 years are able to pass these tests which are termed first-order false belief tests (Wimmer and Perner, 1983). False belief is considered to be the only conclusive test of theory of mind since otherwise the real state of affairs could be appealed to without the need to infer mental states (Dennett 1978). The ability to understand a second-order false belief develops between the ages of 6 and 7 years (Perner and Wimmer, 1985). A second-order false belief is a “belief about a belief” and to pass these tests a child has to
recognise that other people can also represent mental states. Most children with autism cannot solve either first or second order false belief tasks (Baron-Cohen et al. 1985; Perner et al. 1989) irrespective of their mental age or verbal ability. However, some high-functioning individuals can eventually solve first-order false belief tasks, but still usually fail second order tests (Happé, 1993). High-
functioning autism is often referred to as Asperger’s syndrome. These are essentially individuals with autism in which language has developed normally. It is suggested that in solving these first order tests Asperger individuals employ general-purpose reasoning rather than the automatic innate mechanism engaged by normal people.

1.3 Theories of theory of mind

Several types of theory exist which attempt to explain the development of a child’s theory of mind (Carruthers & Smith, 1996; Flavell, 1999). The three most outstanding are the simulation theory (Harris 1992) the theory theory (Gopnik & Wellman, 1994; Perner, 1991) and the modular theory (Leslie, 1987; Leslie, 1994). According to the simulation theory, children use their awareness of their own mental states to infer the mental states of other people through a kind of role-taking or simulation process or in other words imagining oneself to be in the others situation. So with regard to the “Sally-Ann” test, a child can predict where Sally would think her ball is by imagining what they themselves would think if they were in Sally’s shoes and had only the knowledge of where it had been when she had last seen it. Thus, the simulation theory relies upon an innate awareness of one’s own mind. This theory is supported by evidence of the existence of early imitative abilities (Meltzoff and Moore, 1977). In the 1970’s Meltzoff and Moore demonstrated that neonates have an innate ability to imitate a range of facial expressions including tongue and lip protrusion and mouth opening. More
recently, neurophysiologists have discovered specialist neurons (so called “mirror” neurons) in the left inferior frontal gyrus (f5) of the monkey brain which discharge during both performance of the same action and observation of another individual performing an action (Gallese et al., 1996; Gallese and Goldman, 1998). A functional neuroimaging study has found increased activity in the human homologue of monkey area f5, the left inferior frontal cortex (BA44), when an observed action was executed simultaneously (Iacoboni et al., 1999). These results suggest an innate cortical mechanism for human imitation. In the simulation model experience plays an important formative role. It is through practice in role taking that children improve their simulation skills.

However, theory theorists would argue that it is not that we understand the behaviour of others by projecting our own feelings on to them, but that we have an initial representational system which is equally applicable to ourselves and others (Gopnik & Wellman, 1994; Perner, 1991). The theory theory proposes that our understanding of the mind proceeds by the formation, revision and replacement of successive theories of the mind. This theory relies upon the existence of an innate theory formation system that allows us to infer the underlying structure of the world, particularly it’s causal structure, on the basis of events we observe. Also according to the theory theory, there are innate, initial “starting-state” theories about the mind as well as other aspects of the world (Gopnik and Wellman, 1994). These theories are thought to be domain-specific but can be and are radically revised in the light of new evidence, thus they are the
starting state but not the final state of the representational system. Again it is experience that provides children with new information that causes them to revise and improve their current theory of mind.

The modular view of theory of mind is based on our understanding of autism. As autism is the result of an innately determined biological disorder characterised by a specific impairment in theory of mind it follows that there must be an innate basis of theory of mind and thus a specific theory of mind “module”. Leslie (1987; 1994) suggests that theory of mind results from a decoupling of representations of imaginary circumstances from reality. When explaining a person’s behaviour in terms of a belief we have to recognize that this belief might not correspond to reality. Even when it conflicts with reality it is the belief not the reality that determines behaviour. We also have to distinguish the representation of the reality by the other person from our representation of reality. Leslie (Leslie and Thaiss, 1992) has called this the theory of mind mechanism or ToMM. Baron Cohen (1995) has extended this model to give the “metarepresentational system”. He proposed the existence of four sub-components or mechanisms that have evolved to enable the ability to represent mental states; the intentionality detector, eyes-direction detector, the shared attention mechanism and the theory of mind mechanism (ToMM) described by Leslie (Leslie and Thaiss, 1992). According to Baron-Cohen (1995), the four components of the metarepresentational system are domain-specific, with obligatory triggering inputs and can be selectively damaged. The intentionality detector interprets motion stimuli in terms of goals and approach/avoidance desires while the eyes-direction detector interprets visual
stimuli in terms of what an agent sees. These two mechanisms deploy "dyadic" representations, in other words representations which specify the intentional relation between and agent and a state of affairs. The shared attention mechanism is a system dedicated to form "triadic" representations, whose function is to specify the intentional relation among an agent, one self and a third object (e.g. I see – Mummy sees Daddy – ""). The metarepresentational model has the power of bringing both normal and abnormal patterns of development within a single neuropsychological explanatory framework, allowing to make a clear distinction between competence and performance in clinical populations. The modular theory does not entirely exclude either of the other two theories. Indeed it is conceivable that an appropriate theory of theory of mind will incorporate elements from all three of these models.

However, in contrast to these models it has been argued that mental state information processing is closely related to executive functioning (Frye et al., 1995; Frye et al., 1996). Executive functions are processes that underlie purposeful goal-directed actions that contribute to self-regulation. These include set-shifting and set maintenance, interference and attentional control, inhibition, integration across space and time, planning, and working memory (Shallice, 1982). This model has been argued three ways. 1) The development of executive functions allows the child’s theory of mind to develop, or show it’s full potential on theory of mind tasks (Ozonoff et al., 1991; Russell, 1997; Ozonoff, 1997). 2) Theory of mind ability is necessary for the development of executive functioning
(Carruthers, 1996; Perner 1998). 3) There are no specific systems for processing mental states and thus performance on theory of mind tasks can be reduced to executive function ability (Frye et al. 1995; Frye et al., 1996). The motivation for these theories derives from reports of executive dysfunction in autism (Ozonoff et al., 1991; Hughes et al., 1994) and evidence that executive function and theory of mind ability are correlated in normal children (Frye et al., 1995; and Hughes et al., 1998). However, a study by Fine et al. (2001) reported a patient with early or congenital left amygdala damage who demonstrated a profound impairment in theory of mind in the absence of any impairment in executive functioning. This suggests that executive functions do not require the same representational abilities as those involved in mental state processing and supports Leslie' theory (Leslie 1987; Leslie 1994) of a domain-specific neural system. The association in autism of these two impairments may result from the proximity of the respective underlying neuroanatomical systems; executive dysfunction is associated with damage to the pre-frontal cortex (Shallice, 1982), while medial prefrontal cortex has been implicated in theory of mind ability (Goel et al., 1995; Fletcher et al., 1995a; Brunet et al., 2000). Findings of correlations in normally developing children between theory of mind ability and executive functioning may reflect coincidental developmental time-courses between theory of mind and specific executive functions. This supports the notion of the existence of two independent systems for theory of mind and executive functioning but leaves open the possibility that appropriate theory of mind ability is necessary for the development of some executive functions (Carruthers, 1996; Perner 1998).
1.4 Theory of mind deficits

1.4.1 Autism

It is thought that the deficit in autism is not simply an inability to do meta-representation but a specific impairment of mental representations. A number of tests have been developed which compare two behaviours differing only in the demands they make upon the ability to mentalise. This is known as the “fine cuts” technique as it distinguishes between intact and impaired abilities by using cognitively similar tasks. The sabotage/deception task (Sodian and Frith, 1992) is one example of this approach. Other tests include the false photograph/false belief test (Leekham and Perner 1991, Leslie and Thaiss 1992) understanding see/understanding know (Perner et al. 1989), and understanding literal/metaphorical expression (Happe, 1993). Autistic individuals are impaired on each of the tasks involving an understanding of mental states but not it’s cognitive equivalent. Observation of autistic children’s behaviour and conversation has also revealed “fine cut” discrepancies. For instance they spontaneously use instrumental gestures (gestures which are intended to regulate or change the behaviour of others by communicating commands, i.e. go away, be quiet, look over there) but not expressive gestures (gestures deliberately expressing inner feeling states, or responses to feeling states in others i.e. embarrassment, consolation, anger) (Attwood et al. 1988). As with most higher cognitive functions, mentalising is a process that is not easily isolated. It is a
highly sophisticated and ubiquitous element of social reasoning and problem solving. The “fine cuts” approach is thus considered to be well suited for neuroimaging studies of mentalising, in both normalise and patients, using “cognitive subtraction”.

1.4.2 Acquired deficits

It is not however just autistic individuals who display a lack of “theory of mind”. Some patients with schizophrenia demonstrate post-morbid impairments in mentalising capacity (Frith and Corcoran, 1996; Corcoran et al., 1997; Pickup and Frith, 2001), as do some patients with right hemisphere damage (Happé et al. 1999) and frontal lobe damage (Channon and Crawford, 2000; Stuss et al. 2001; Rowe et al. 2001). These disorders are in contrast to autism in that the deficit is acquired. In Autism the deficit in “theory of mind” probably results from a malfunction early on in the development of either the brain regions directly involved in mentalising or auxilliary brain networks/regions required for appropriate development of this capacity. The fact that this ability can be “taken out” has implications for the brain basis of theory of mind, supporting the notion of a dedicated cognitive system. Renewed interest in the neural substrates of “theory of mind” and it’s implications for the biological basis of autism has engendered a growing body of evidence from neuroimaging studies of both patients and normal volunteers of this highly specialised and very human ability.
1.5 Neuroimaging and theory of mind

1.5.1 Neuroanatomical Imaging

Morphological examinations of the brain provide cognitive neuroscientists with important insights into its functional organisation. In the past such examinations had to be performed post mortem, thus limiting the amount of data available. Advances in MR imaging techniques have provided a high resolution instrument for measuring the structures of the cognitively impaired brain, invivo. The main advantages of invivo studies are threefold; 1. Patients can be assessed at the time of examination and an accurate measure of their impairment can be established. 2. Longitudinal studies inform on the progress or recovery of a disease 3. Patients can be studied at the onset of symptoms in a drug naïve state. The study of autism in this manner may reveal the brain abnormalities of this disorder and go some way to representing the biological basis of the “social brain”.

Abell et al. (1999) performed a voxel-based whole brain analysis of structural MRI data from 15 high functioning individuals with Asperger syndrome compared with 15 normal volunteers matched on sex, handedness, age and performance on tests of verbal and non-verbal ability. Theory of mind ability was tested in both groups; all subjects were able to pass the first-order false belief tasks, but performance on the two second-order false belief tasks was significantly poorer in the Asperger group of whom only four were able to pass both tests. The comparisons from the autistic group showed decreases in grey matter volume in
the right paracingulate cortex, the left occipito-temporal cortex and the left inferior frontal sulcus and increases in the left amygdala/peri-amygdaloid cortex, the right inferior temporal gyrus, the left middle temporal gyrus and parts of the cerebellum. A more recent study by Howard et al., (2000) investigated the "absolute" volume of specific brain structures in people with high functioning autism (HFA). The areas measured were temporal and medial temporal structures, lateral ventricles, right and left cerebral hemispheres and intercranial volume. They found (consistent with Abell et al., 1999) bilaterally enlarged amygdala volumes in the HFA group compared to a control group.

The areas of abnormal grey matter volume seen by Abell et al. (1999) in the autistic group (with the exception of the cerebellum) form part of an amygdala centered circuit (Amaral et al., 1992). The ventral temporal cortex (BA 20 and 21) sends connections to the amygdala which in turn has reciprocal projections to the inferior prefrontal convexity (BA 12/45) and anterior cingulate cortex (BA 24 and 32). This circuit has been implicated in emotional and social learning by animal and human neuropsychological studies (for review see Kling and Brothers, 1992). One theoretical view of brain plasticity proposes that the amygdala plays a central role in integrating highly processed perceptual inputs that have value and salience and uses this synthesis to adapt behaviour by modulating synaptic changes in the relevant regions of the circuit. (Friston et al., 1994). This model suggests that a neurodevelopmental abnormality involving the amygdala or its targets would be sufficient to explain the impairments in emotional and social
learning in autism, and can account for morphological abnormalities at the sources of amygdala inputs. These results do not speak directly to the neural correlates of the theory of mind mechanism, however, if the notion of a dedicated system is correct they suggest that such a system is a component of this amygdala centered network. It is still necessary to determine the contribution of the individual regions in this network during mentalising. In order to achieve this we need to look at how these regions interact while processing theory of mind tasks.

1.5.2 Functional Imaging

A number of studies using tasks which have been developed and validated with autistic people have sought to examine the neural substrates of mentalising ability. Using PET, Fletcher et al. (1995a) scanned volunteers whilst reading and answering questions about short stories (Fig 3). In one type of story (theory of mind stories) the behaviour of the protagonists could only be explained on the basis of their (false) beliefs. In a second type of story the events could be understood on the basis of physical causality or general knowledge. In this instance, the beliefs and intentions of the characters were irrelevant. A third condition consisted of unlinked sentences. This condition acted as a control for reading, attention to sentence meaning, memory and question answering. Comparison of activation during “theory of mind” stories versus the “physical causality” stories revealed task-specific activation in the left medial prefrontal gyrus (predominantly BA8) (x=-12, y=36, z=36) and the anterior cingulate cortex, as well as increased activation of the posterior cingulate cortex and the inferior
parietal lobe on the right. However, only the left medial prefrontal activation was found to be unique to mentalising. Happé et al. (1996) repeated this study on autistic individuals diagnosed with Asperger's syndrome. Scanning patients who are unable to perform the task is meaningless, however a small minority of autistics are able to pass standard false belief tests. It is believed that this ability does not originate from the same automatic and innate mechanism as normals, but may reflect a general purpose reasoning and learned process. This is highlighted by their poorer performance on more naturalistic mentalising tests (Happé, 1993).

The pattern of activity was the same as in the Fletcher et al. (1995a) study with one exception, although differences in activation between conditions was smaller in the Asperger group. The exceptional activation was in the medial pre-frontal region. The Aspergers' group activated a sub-region of the same medial prefrontal region as the normal volunteers. This sub-region was more ventral (BA 9/10) to the maximal activation of the previous study. BA 9 has been implicated in imaging studies of general cognitive ability and particularly problem solving (Dolan et al. 1992) which fits with the idea that Asperger individuals apply a general purpose reasoning mechanism when tackling theory of mind tests.

Brunet et al., (2000) examined "theory of mind" using a non-verbal comic strip adaptation of the story comprehension paradigm. One condition required subjects to attribute intentions to the characters of the comic strips. The other two conditions involved only physical logic and knowledge about objects' properties; one involved characters whereas the other only represented objects. They found
widespread activations to be associated with attribution of intentions. In particularly bilateral activations in the anterior cingulate cortex and the right medial prefrontal cortex \((x=4, y=56, z=44)\) (anterior paracingulate cortex, BA 32). Interestingly, they also found bilateral temporal activation to be associated with the involvement of characters but not specifically mentalising.

A similar pattern of activation has been found in another study (Castelli et al., 2000) in which volunteers were scanned in PET while watching silent computer-presented animations. The characters in the animations were two simple geometrical shapes whose movement patterns selectively evoked mental state attribution or simple action description either goal directed or random, based entirely on their kinematic properties. The mental state attribution condition was compared with the random action and again the medial prefrontal cortex was activated, only slightly more anteriorly than in previous studies \((x=-6, y=58, z=32)\). A number of other regions also activated including the right temporo-parietal junction basal temporal cortex and the occipital cortex. The goal directed condition elicited intermediary activity in the same network of regions when compared to the random sequence. The above results suggest that these regions form a network for processing information about intentions driving behaviour.

The medial prefrontal activation seen by both Fletcher et al. (1995a) and Brunet et al. (2000) was also seen in another imaging study of “theory of mind”. Goel et al. (1995) used PET to scan adults engaged in a complex mentalising task in which
subjects were shown pictures of familiar, modern day and unfamiliar, pre-fifteenth century, man-made artifacts. The authors used a hierarchical paradigm which built upon the cognitive processes used in the baseline condition for three other conditions, culminating in a theory of mind task. The baseline task was believed to require visual perception, shape discrimination, simple decision making and motor response. The second condition also required memory retrieval, and the third required inference of the functional information based on pictorial cues such as shape, texture, material and size. The "theory of mind" condition incorporated all the previous processes but also required the subject to model the knowledge and rationality of another mind. They found widespread activation associated with this task, including activation of left medial frontal lobe and left temporal lobe. The increasing complexity of the task in this study may account for the medial frontal activation, consistent with the findings of Paus et al. (1998) which concluded that task difficulty plays a major role in modulating blood-flow response in the anterior cingulate cortex. However, the authors (Goel et al., 1995) would dispute this as the region was not active during the memory and the inference conditions which were confirmed to be more difficult than the baseline, verified by increasing reaction times. They also claim this region to be anterior to the anterior cingulate, in Brodmann’s area 9. Unfortunately, the low spatial resolution of the PET technique means that precise localisation of this region can not be achieved and therefore anterior cingulate activation cannot be ruled out. Fletcher et al. (1995a) excluded task difficulty as a possible confound for their study by matching the "physical" and "theory of mind" conditions for all
but the mentalising factor. Indeed, it was found that the physical stories were read significantly slower than the theory of mind stories suggesting that these were, in fact, the more difficult tests.

1.6 \textit{Functions of the anterior paracingulate cortex}

The PET functional imaging studies mentioned above have all activated a region of the medial prefrontal cortex corresponding to either areas 8/32 or 9/32. This region is known as the anterior paracingulate cortex. This is one of the regions that Abell et al. (1999) found to have decreased grey matter in Asperger's patients. The anterior cingulate cortex (ACC) has long been considered as belonging to the limbic system. It has classically been related to affect. However, recent studies have revealed the cingulate to be both functionally and cytoarchitecturally heterogeneous, comprising numerous specialised subdivisions that subserve an array of cognitive, emotional, motor, nociceptive and visuospatial functions (Vogt et al., 1992). A review of imaging studies by Bush et al. (2000) has shown that separate regions of the ACC are involved in cognition (dorsal ACC) and emotion (rostral ACC). The areas of the rostral anterior cingulate involved in processing emotion are BA 24, 25 and 33. The area activated in previous studies of mentalising ability is anterior to these regions, in the paracingulate cortex (BA 32) and medial prefrontal cortex (BA 9/10). The paracingulate (BA32) shares similar connections to the ACC - amygdala, hippocampus, hypothalamus, anterior insula and orbito-frontal cortex - but is
cytoarchitectonically distinct and has been described as a cingulo-frontal transition area (Devinsky et al., 1995). The specific function of the region of the anterior paracingulate cortex demonstrated in mentalising studies, is still largely unknown. However this region is clearly anterior to the motor region of the anterior cingulate as described by Paus et al. (1998). Adjacent regions of the paracingulate cortex have activated in studies which have involved some mode of self-monitoring; visual self-recognition (Kircher et al. 2000), memory for autobiographical events (Maguire and Mummery, 1999; Maguire et al. 2000), verbal self-monitoring (McGuire et al. 1996a), self-generated thoughts (McGuire et al. 1996b), externally-produced tickling (Blakemore et al. 1998), perception of pain (Rainville et al. 1997), reporting emotional response (Lane et al. 1997) and monitoring intended response (Carter et al. 1998). Additionally, medial prefrontal areas have been shown to activate during metaphor comprehension (Bottini et al., 1994), which according to some theorists requires recognition of the speakers intentions. Supporting evidence from the neuropsychological literature has shown that the frontal lobes are necessary for inferring the mental states of others (Channon and Crawford, 2000; Stuss et al. 2001; Rowe et al. 2001). In particular patients with bi-lateral medial frontal lesions have been shown to be impaired on tasks of deception (Stuss et al. 2001). Taken together, these results demonstrate that the cognitive mechanism behind the ability to monitor one’s own mind and the minds of others is mediated by this highly circumscribed region of the medial prefrontal cortex, regardless of the specific source of information.

1.7 The amygdala and theory of mind
It is clear that mentalising recruits regions of the brain involved in the general interpretation of behaviour which contribute to a neural network for social cognition, the extent of which remains to be identified and the specific functions, to be determined. Several other brain regions have been implicated in theory of mind which may contribute to this network, in particular the amygdala (Baron-Cohen et al. 1999). Fine et al. (2001) reported a single case study of a patient with early or congenital left amygdala damage. They found their patient to be severely impaired in his ability to represent mental states, yet he showed normal executive functioning. Baron-Cohen et al. (1999) reported activation in the left amygdala (in addition to medial prefrontal activation) during a task in which normal volunteers were required to infer the mental state of an individual from the expression of their eyes. They also found significantly less amygdala activation when adults with high functioning autism and Asperger syndrome performed the same task. Neuroanatomical studies of autism; post-mortem (Bauman and Kemper, 1994; Courchesne, 1997) and in-vivo (Abell et al. 1999; Howard et al., 2000) have found abnormalities in the amygdala complex. Neuroimaging studies have demonstrated a role for the amygdala in processing facial expressions (Morris et al. 1996), monitoring gaze (Kawashima et al. 1999; Wicker et al 1998) and the perception of biological motion (Bonda et al., 1996). All of which are functions with important social significance. Thus we may speculate that the amygdala may be recruited during theory of mind processing to provide additional social information or that normal amygdala function may be a prerequisite for the
development of theory of mind. In other words, normal amygdala function in young children may be essential for formulating representations of mental states in the anterior paracingulate cortex. Thus, we would expect that as the child's theory of mind ability matures, the amygdala is recruited less and less. This would explain why lesions to the medial prefrontal lobe with no amygdala involvement are sufficient to "take-out" a pre-morbid functioning mentalising ability in later life (Stuss et al. 2001) and why many studies of theory of mind in adults do not activate the amygdala.

1.8 Questions explored in this thesis

In this thesis a series of experiments investigating further the neural correlates of theory of mind will be presented. The first study sought to examine anatomical convergence between verbal and non-verbal mentalising tasks by exploiting the superior spatial resolution of fMRI over PET. The story comprehension task used in the Fletcher et al. (1995a) study was adapted for compatibility with fMRI. This was compared with an equivalent non-verbal task involving the meaning of captionless cartoons. Previous behavioural studies have used both stories and visual jokes to investigate theory of mind impairments in adults and children and found these good markers of mentalising abilities.

The aim of the second experiment was to devise an "online" mentalising task to examine mentalising in real time. Previous studies have used tasks in which the
subject is presented with a scenario and they have to explain why the persons behaved the way they did. The “online” mentalising task, which was a modification of the playground game “stone paper scissors”, required subjects to predict the response of their opponent and outwit or second guess them.

The aim of the final three studies was to explore the effects of theory of mind deficits and normal function on one aspect of everyday communication; the recognition of gesture. A previous study has shown that Autistic children do not use gestures to express their feelings (expressive) but do use gestures which convey commands (instrumental) (Attwood et al. 1988). This is thought to be due to their impaired theory of mind ability. Chapter 5 describes an initial behavioural study conducted to ascertain the developmental age at which normal children are able to represent gestures and expose differences between instrumental and expressive gestures. The same paradigm was then tested on Autistic children, to confirm that the non-use of expressive gestures is due to a lack of understanding, resulting from their inability to represent the emotions of others rather than an effect of the paradigm itself. Chapter 6 describes an imaging study performed on normal adults to establish the neural substrates of representing gestures and examine differences between expressive and instrumental gestures. In this study normal adult volunteers were imaged while just observing or observing and recognising the gestures presented in short videos.
The results from these studies are discussed in relation to the cognitive mechanisms underpinning our everyday ability to 'mind-read', with reference to existing models and theories of the evolution and development of this ability and their implications for our better understanding of the biological basis of autism.

1.9 Acknowledgement of assistance

The majority of the work involved in the production of this thesis was performed by the author herself. This includes the building of paradigms, preparation of stimuli, administration of pilot studies, recruitment, consent and briefing/debriefing of volunteers, the imaging (both coordination of stimulus and technical running of the scanner), data collection, analysis and interpretation of results. However, there are a number of people who have contributed stimuli or expertise, whom I would like to acknowledge and thank here. The stories in chapter 3 were provided by Paul Fletcher and Francesca Happé and were used previously in their respective studies (Fletcher et al., 1995; Happé et al., 1996). Also, some of the cartoons were supplied by Francesca. The program used in the administration of the stimulus in chapter 4 was written and supplied specifically for this study by Anthony Jack. Finally, the camera work in the videos in chapters 5 and 6 was performed by Peter Aston who also acted in half of the videos in chapter 6.
Chapter 2: METHODS, ANALYSIS AND EXPERIMENTAL DESIGN IN FUNCTIONAL NEUROIMAGING

2.1 Introduction

The work presented in this thesis uses the biological imaging techniques of Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI) to investigate neuronal response to behavioural stimuli. These techniques allow the direct measurement of components of the cerebral haemodynamic response that can be used as an indirect reference for neural activity. This chapter explains the basic principles behind the two imaging techniques, experimental design and statistical analysis.

2.2 Functional neuroimaging methods

2.2.1 The relationship between neuronal activity and cerebral blood flow

The fundamental physiological principle of functional neuroimaging is represented by the tight coupling which exists between neuronal activity and the associated local increase in blood flow and energy metabolism (Raichle, 1994). This was first proposed by Roy and Sherrington (1890) and has been supported by many studies since (for review see Magistretti and Pellerin, 1999). When a population of neurons becomes activated it requires an increased supply of blood flowing to it to replenish the supply of oxygen and glucose, which the neurons
depend on for energy. Thus, functional neuroimaging techniques measure neuronal activity by detecting changes in blood flow or flow related phenomena in humans. There are a number of functional neuroimaging techniques; the two used in this thesis are PET and fMRI.

2.2.2 Positron Emission Tomography (PET)

2.2.2.1 General overview

Positron emission tomography (PET) is a technique for measuring the concentrations of positron-emitting radioisotopes within the tissue of living subjects. Positron-emitting isotopes are produced by a cyclotron and used to label molecules of a compound of interest whose distribution can then be examined directly. A positron is an anti-particle of an electron with the same mass but opposite charge. The unstable nucleus of a positron-emitting isotope emits both a positron and a neutrino. The positron has some initial kinetic energy that is lost through collisions with neighbouring electrons in surrounding tissue. When the positron loses all or most of its energy it eventually recombines with an electron to form a positronium. Almost instantaneously the positronium annihilates, forming two gamma-rays (γ) with equal energy emitted at 180° to each other, to conserve energy and momentum. The energy of these γ-rays is sufficient to escape the body and can be detected by a PET camera. The PET camera consists of circumferential arrays of scintillation detectors that identify emission of γ-rays. When two opposing detectors produce a signal within a narrow time interval (15ηs) a coincidence event has been detected. Since annihilation photons are
emitted simultaneously and in opposite directions the detection of a coincidence event localises the positron annihilation to a point lying somewhere on the line that joins the two detectors. By combining data from many detectors simultaneously and using a back projection algorithm the count density within the head can be determined.

The study described in chapter 4 used a single-positron emitting substance $^2H_2^{15}O$ (radioactive water) introduced into the bloodstream via a bolus intravenous injection. This tracer diffuses freely across the blood brain barrier and so uptake in cerebral tissue is determined by the blood flow to that tissue rather than the diffusion rate of the tracer into the tissue. After injection of the bolus of $^{2}H_2^{15}O$, the tracer is extracted from plasma into cerebral tissue on the first pass through the brain and the tracer uptake is highly correlated with cerebral blood flow. $^{2}H_2^{15}O$ has a half-life of approximately 2 minutes and therefore repeated measures every 8 minutes are possible. Water activation studies measure changes in regional cerebral blood flow (rCBF) as a result of cognitive or sensorimotor stimulation.

2.2.2.2 PET experimental design

The PET study described in this thesis was performed using an ECAT EXACT HR+ scanning system (CTI Siemens, Knoxville, TN) in 3D mode with septa retracted (Townsend, 1991) with increased sensitivity to head counts produced by the use of a NeuroInsert (Grootoont et al., 1996). $^{2}H_2^{15}O$ (radioactive water) was
administered through a venous cannula inserted into the antecubital fossa vein of the volunteer. Approximately 350 MBq of H$_2^{15}$O in 3ml of normal saline was loaded into intravenous tubing and flushed into subjects over 20s at a rate of 10ml/min by an automatic pump. After a delay of approximately 35s a rise in counts could be detected in the head that peaked 30-40s later (depending on individual circulation time; Silbersweig et al., 1993). The interval between successive H$_2^{15}$O administrations was 8 minutes. The data were acquired in one 90s frame, beginning during the rising phase of the head curve. Correction for attenuation was made using a transmission scan collected at the beginning of each study. Images were reconstructed by filtered back projection (Hanning filter, cut off frequency 0.5 cycles per pixel) into 63 image planes (separation 2.4mm) and into a 128 x 128 pixel image matrix (size 2.1mm). Each volunteer received 9mCi of activity per scan, and twelve scans were acquired per volunteer in each experiment, giving a total of 108mCi of radioactivity. Stimuli were sequences of static images presented in epochs of 90 seconds, the orders of the tasks and conditions were counterbalanced across volunteers. The duration of the scanning in each experiment was approximately 2 hours. Details of individual methodologies are described in the appropriate chapters.
2.2.3 Functional Magnetic Resonance Imaging (fMRI)

2.2.3.1 General Overview

Magnetic resonance imaging (MRI) is based on the spinning motion of a specific kind of nuclei present in biological tissues – these are MR active nuclei. MR active nuclei are characterised by their tendency to align their axis of rotation to an applied magnetic field. The hydrogen nucleus is the MR active nucleus used in MRI. The hydrogen nucleus contains a single photon (atomic number 1). It is used because it is very abundant in the human body, and because its solitary proton gives it a large magnetic moment. This is the vector sum of the individual magnetic moments of all the protons in the nucleus and is the physical property responsible for the alignment in the large static magnetic field of the MRI scanner. By displacing the alignment of the magnetic moments through excitation with a radiofrequency pulse, and then measuring the energy given up by the nuclei to the surrounding environment as they recover their longitudinal magnetic moment, an image that reflects the tissue environment of the protons can be produced. The rate of recovery is an exponential process, with a recovery time constant called T1. This is the time it takes 63% of the longitudinal magnetisation to recover in the tissue. T1 recovery is one of three sources of contrast; T1 recovery, T2 decay, T2* decay and proton density, in the MR image. T2 decay is caused by nuclei exchanging energy with neighbouring nuclei and results in a decay or loss of transverse magnetisation (magnetisation in the transverse plane) caused by precession of the spinning nuclei about the longitudinal axis of the
static magnetic field. The rate of decay is also an exponential process, so that the
T2 relaxation time of a tissue is its time constant of decay. It is the time it takes
63% of the transverse magnetisation to be lost.

Individual biological tissues have different T1 and T2 constants due to their
molecular structure. Fat has a shorter T1 and T2 than water. T2* decay is faster
than T2 decay since it is a combination of two effects, T2 decay itself and
dephasing due to magnetic field inhomogeneities. These are areas of the magnetic
field that do not exactly match the external magnetic field strength and are caused
by the presence of paramagnetic substances in the tissues. The proton density of a
tissue is the number of protons per unit volume of that tissue. The resonance
frequencies of the spinning nuclei are proportional to the magnetic field. By
applying a linear magnetic field gradient it is possible to change the resonance
frequencies of the spins. In this way the location and quantity of the spins along
the gradient axis can be determined and used to form an image. Because there is
not much variation in spin density over different tissues the contrast between them
will not be very large. The differences in the tissues T1 and T2 can be used to
increase image contrast. By choosing appropriate sequence parameters the scan
can be “weighted” to T1, T2, T2* or proton density differences between the
imaged environments. This is known as T1-, T2-, T2* or proton density-
weighting.
In the early 1990's it was realised that variations in T2* could be exploited to measure cerebral blood flow via injection of a paramagnetic vascular marker such as gadolinium. The combination of this technique with an ultrafast image acquisition technique, echoplanar imaging (EPI) (Mansfield, 1977), allowed dynamic imaging of the passage of such a contrast agent through the brain. This led to the first measurements of changes in cortical perfusion with MRI (Belliveau et al., 1991), that could be thought of as loosely analogous to a PET measurement. However, since these early experiments the endogenous contrast properties of deoxyhaemoglobin in the blood have now been utilised for a completely non-invasive fMRI procedure. Deoxyhaemoglobin is more paramagnetic than oxyhaemoglobin; thus haemodynamic changes that alter the relative proportions of deoxyhaemoglobin to oxyhaemoglobin will change the magnetic environment and therefore the MR signal around these cortical vessels. The more paramagnetic an agent the faster its relaxation time is, and the less signal it will produce. Therefore deoxyhaemoglobin produces less MRI signal than oxyhaemoglobin. This was originally demonstrated by Ogawa et al., 1990, who observed blood oxygenation level dependent (BOLD) contrast in anaesthetised mice at high magnetic field strengths (7Tesla and 8.4Tesla). In T2* weighted gradient echo images they observed less signal in brain scans when mice were breathing 20% oxygen than when they were breathing 100% oxygen. These signal changes were reversible. The authors suggested that deoxyhaemoglobin was the source of the contrast, enhancing relaxation in and around the vasculature, resulting in the observed loss of signal. The development of EPI sequences specifically sensitive
to T2* allowed this non-invasive functional imaging of changes in cortical haemodynamics related to cerebral activity (Kwong et al., 1992; Ogawa et al., 1993). This technique is used in the fMRI experiments in this thesis.

The exact relationship between neural activity and BOLD contrast is still under investigation (Toga and Mazziotta, 1996), but the basic principles have been established (Malonek and Grinvald, 1996). Neural activity in the brain appears to cause an initial increase in oxygen consumption, leading to a rise in deoxyhaemoglobin concentration and a fall in the MR signal that can be seen at high field strengths (Menon et al., 1995). This brief ‘early dip’ is then superseded by a large rise in the signal that is thought to reflect a transient increase in blood flow to the active area. Because the flow increase temporarily outweighs the consumption, a transient decrease in deoxyhaemoglobin concentration occurs, giving rise to a large peak in the MRI signal that is easily seen using EPI at field strengths of 1.5T and greater. At the lower field strengths (≤ 2 Tesla) used in this thesis, this latter signal is dominant and the ‘early dip’ is not usually seen.

Kwong et al. (1992) and Ogawa et al. (1992) reported a change in MRI signal resulting from a change in concentration of deoxyhaemoglobin in human visual cortex. Both groups scanned volunteers using BOLD fMRI while viewing bright light (the experimental task) or at rest (the baseline). By subtracting rest state images from activation images, Kwong demonstrated activity-induced changes during the visual task. The signal from the visual cortical areas followed the on-
Methods

Chapter 2

off activation paradigm of the visual task, with a lag of several seconds between the visual stimulation and BOLD signal. This reflects the delay in vascular and metabolic responses and is known as the haemodynamic response time. Furthermore, the BOLD changes in the visual cortex occurred primarily in the grey matter, not in the white matter, as would be expected (Ogawa et al. 1992).

The rise in BOLD signal during visual stimulation indicates that there is a decrease in the concentration of paramagnetic deoxyhaemoglobin during stimulation relative to rest. Earlier research using PET demonstrated that there is a disproportionate increase in the amount of oxygenated blood flowing to an activated region, in that the rise in oxygen re-uptake during somatosensory and visual stimuli is smaller than the rise in blood flow to activated brain regions (Fox and Raichle, 1986). Earlier observations during open skull surgery also demonstrated that blood leaving an active cortical region is brighter red, i.e more oxygenated, than when the region is not active, as a result of this mismatch between demand and supply. (Penfield, 1933).

2.2.3.2 fMRI Experimental design

The studies described in chapters 3 and 7 were performed on a Siemens VISION MRI system (Siemens, Erlangen) operating at 2 Tesla, using echo-planar T2* weighted image volumes with blood oxygenation level-dependent (BOLD) contrast. Each study involved scanning between six and twelve subjects. Each image volume constituted either 48 contiguous 3mm axial slices (chapter 3) or 36 contiguous 3mm axial slices (chapter 7) with in-plane resolution of 3mm by 3mm.
positioned to cover the whole brain. fMRI acquisition was continuous with each image volume acquired every 2-4 seconds. Stimuli were either static images or short videos presented in short epochs, the orders of the tasks and conditions were counterbalanced across volunteers. Activation epochs were interspersed with control (rest) conditions. Each run began with six volumes which were discarded prior to analysis to allow for T1 saturation effects. The duration of the scanning in each experiment was approximately 40 to 45 minutes. Details of individual methodologies are described in the appropriate chapters.

2.2.4 Structural Magnetic Resonance Imaging

A high-resolution structural image of the brain was acquired on the same Siemens VISION MRI system (Siemens, Erlangen) for each volunteer. This allows anatomical co-registration with the functional data from both PET and fMRI studies. The sequence used was a modified magnetisation prepared rapidly acquired gradient echo (MP-RAGE) (Deichmann et al., 2000). The advantages of using 3D imaging, where the entire volume is stimulated and not the individual slices, are an increased signal to noise ratio (SNR), thin contiguous slices and the facility for using multi-planar reconstruction. The resultant resolution is: in-plane 1mm x 1 mm and slice thickness 1mm.

2.2.5 Comparing PET and fMRI

Although PET and fMRI are both useful tools for imaging brain function, there are important differences between the two techniques that need to be considered when deciding which of them is most appropriate for an experiment. The most
important are the differences in spatial and temporal resolution. PET measures blood flow on a spatio-temporal scale of about 10-15 mm and 30 seconds. PET measures changes in blood flow or perfusion directly in terms of the amount of radiolabelled water that accumulates locally but this takes several tens of seconds to a minute. fMRI has a spatio-temporal scale of about 1-3mm and one or more seconds. For the fMRI sequences used in this thesis the effective spatial resolution is 5-10mm. The lower limits on the effective resolution of fMRI are physiological and imposed by the spatio-temporal organisation of evoked haemodynamic responses (2-5mm and 5-8 seconds). Local increases in neural activity cause both a relative deoxygenation of blood and an increase in perfusion, which quickly reverses the deoxygenation, leading to an increase in oxygenation that endures several seconds.

fMRI is non-invasive and unlike PET does not involve the use of potentially harmful ionising radiation. This also means that volunteers can return for multiple sessions allowing longitudinal studies. The superior temporal resolution of fMRI opens the way to ‘event-related’ designs (as well as epoch or block designs), where the modulation of the haemodynamic response by brief events is detected, whereas PET is limited to epoch designs. Nonetheless, there are important technical difficulties with fMRI that do not always make it the technique of choice. While PET is completely silent, fMRI produces noise levels of around 100dB in the bore, thus causing restrictions on studies where perceptual load is manipulated in the auditory domain. Also, unlike PET, fMRI is particularly
susceptible to motion artefact. Thus studies with a motor component or involving overt speech may be confounded. In PET, access to the volunteer is unrestricted if additional apparatus is required; however in fMRI the volunteer is entirely enclosed in the RF coil and bore of the magnet. This leads to a restricted field of view (compared to PET) for the presentation of visual stimuli, and difficulty introducing any other instrumentation. Despite the superior spatial resolution of fMRI, the technique has important spatial limitations that affect the ability to image certain areas of the brain. Near air-filled sinuses (anterior and inferior frontal regions or mid temporal lobes), artefacts and signal drop out due to dephasing of spins typically mean that these areas are not imaged well with BOLD fMRI. These difficulties vary considerably between subjects, depending on their individual anatomy.

However these have not been important issues in the work presented here. Thus fMRI has been the technique of choice, with the exception of chapter 4 which used PET. Low frequency drift has often been reported in time series fMRI data, which has been attributed to scanner instabilities (Smith et al., 1999). Thus, stimuli should be presented in epochs of no more than 30 seconds. In the PET experiment of chapter 4 a long lead-in of the stimuli (approximately 15 trials, each 7s duration) was necessary before data acquisition. In this circumstance the block design of PET was deemed more appropriate.
2.3 Experimental designs

2.3.1 Functional segregation and functional integration

The main aim of functional neuroimaging is to localise cognitive, sensory or motor functions in the cortex of the brain based on the two models of functional organisation to which the brain seemingly adheres; functional segregation and functional integration. Functional segregation demands that cells with common functional properties be grouped together (Zeki, 1990). Zeki proposed that this architectural constraint necessitates cortical connections to converge into and diverge from any particular brain region. Connections between cortical regions are not continuous but occur in patches or clusters and in some areas, in particular the visual cortex, this patchiness has a clear relationship to functional segregation and specialisation. If neurons in a given cortical area share the same responsiveness to some sensorimotor or cognitive process then their segregation is functional as well as anatomical. The data analysis used in the studies in this thesis examines functional neuroimaging data for regionally specific effects based on this model of cortical functional and anatomical segregation. Scanning a volunteer whilst they perform a cognitive or sensorimotor task should lead to activity changes in, and only in, the region of interest.

Functional integration is mediated by the interactions between functionally segregated areas. These interactions can be characterised in terms of functional or effective connectivity, which is the influence that one neural system exerts over
another (Buchel and Friston, 1997). Functional integration is usually assessed by examining the correlations among activity in different brain regions, or by attempting to explain the activity in one area in relation to activities in other brain areas.

A number of experimental designs are in common use with studies of functional neuroimaging such as cognitive subtraction, factorial designs, parametric designs or psychophysiological interactions. The following describes the two types of experimental design used in this thesis; subtraction and factorial.

2.3.2 Subtraction designs
Cognitive subtraction is the simplest and most widely used design in functional neuroimaging. In this approach, the neural correlates of a particular cognitive or sensorimotor process can be investigated by using two or more tasks that differ by one separable component. Regionally specific differences in brain activity between the tasks can then be attributed to this component of interest. Typically volunteers are scanned while performing an experimental task that engages the component of interest and one or more baseline tasks which engage all the components of the experimental task with the exception of the component of interest. Subtraction has been successfully applied to studies of functional specialisation in the extrastriate cortex (Lueck et al., 1989) and the functional anatomy of word processing (Petersen et al., 1989). Unfortunately this simplistic approach makes the assumption that cognitive states differ in components that can
be purely inserted or removed with no interaction between them, both at the level of a function and at the level of its neural implementation. When a component is added to a task’s requirements, the new task comprises not only the previous components and the new component, but also the integration of the new and old components. This integration or interaction could also be considered as an additional component.

One way of eliminating interaction effects is to use a cognitive conjunction analysis. Cognitive conjunctions combine a series of subtractions with the aim of isolating a cognitive component of interest that is common to two or more task pairs. Here several hypotheses are tested, asking whether all the activations in a series of task pairs, are jointly significant. In this way the assumption of pure insertion in subtractive designs can be avoided by extracting the presence of a main effect in the absence of an interaction.

2.3.3 Factorial designs

In some circumstances the interactions between two components may be of interest. An interaction is basically a change in a change. Interactions are associated with factorial designs where two factors are combined in the same experiment. Two factors interact if the level of one factor affects the level of another. In functional neuroimaging, the main neural effects of each factor and the neural effects of the interaction between the factors can be calculated separately. The interaction identifies brain areas where the effect of one factor varies
depending on the presence or absence of the other. This allows the effect of one factor on the expression of the other factor to be measured explicitly (Figure 2-1).

### Figure 2-1

<table>
<thead>
<tr>
<th></th>
<th>With Factor A</th>
<th>Without Factor A</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Factor B</td>
<td>Condition 1</td>
<td>Condition 2</td>
</tr>
<tr>
<td>Without Factor B</td>
<td>Condition 3</td>
<td>Condition 4</td>
</tr>
</tbody>
</table>

Interaction effect (A x B)

In factorial designs there are two or more factors, in this case factors A and B. The main effects of each factor can be calculated by $(1+2)-(3+4)$ and $(1+3)-(2+4)$. In addition, the interaction between the factors can be calculated separately by $(1-2)-(3-4)$ and conversely by $(3-4)-(1-2)$. The interaction identifies areas where the effect of one factor varies depending on the presence or absence of the other.

### 2.3 Statistical analysis

#### 2.3.1 Background

Functional images acquired in both PET and fMRI must undergo rigorous pre-processing before changes in haemodynamic activity produced by experimental conditions can be estimated. These include realignment of the image series, coregistration of the low resolution functional images into the same stereotactic space as the high resolution structural images, spatial normalisation and smoothing. In this thesis a statistical package, Statistical Parametric Mapping

56
(SPM) (Wellcome Department of Imaging Neuroscience, http://www.fil.ion.ucl.ac.uk) custom written for functional neuroimaging was used to perform these steps.

2.4.2 Realignment

Realignment corrects for motion of the volunteer during the course of data acquisition. In PET twelve image volumes are acquired per volunteer in each study. These are acquired over 2 hours. In fMRI the total acquisition time is typically less, of the order of 10 to 20 minutes, but the number of image volumes is significantly greater. Movement, however small, is inevitable. Image volumes are realigned to the first in the series. A least squares algorithm (Friston et al., 1995a) corrects for the three affine transformations (x, y and z) and the three rotations, (pitch, roll and yaw). In practice movements of the order of a few millimeters only are seen. Excessive movement can lead to false activations at the statistical inference stage. This is particularly true of fMRI where head motion can result in signal artefact resembling BOLD activations. This artefactual signal is especially a problem when the head motion is correlated with the stimulus, which can occur in paradigms that involve movement of a limb or overt speech either as a task or as a response to a task (Hajnal et al., 1994).

After realignment the functional images are co-registered to the high resolution T1 weighted structural volume, thus putting all the acquired images from a single volunteer into the same stereotactic anatomical space.
2.4.3 Spatial normalisation and smoothing

To facilitate between subject comparisons and reporting of data in a stereotactic reference the realigned and co-registered images are spatially transformed to conform approximately to a standard brain. This is known as normalisation. It involves determining a twelve parameter transformation between one of the experimentally acquired images and a template in stereotactic space (Friston et al., 1995a). These parameters are then applied to all the functional and structural images, bringing them into standard anatomical space. In this thesis the stereotactic space of the Montreal Neurological Institute, which conforms to that of the atlas of Talairach and Tournoux (1988) was used. The normalisation parameters are determined from a T1 structural image that has been co-registered with the functional images, or a mean functional image (PET or fMRI) to the appropriate template (rCBF or EPI T2*).

Following normalisation, the images are smoothed or convolved with a gaussian kernel. This increases the signal to noise ratio as the neurophysiological effects of interest are produced by haemodynamic changes expressed over spatial scales of a few millimeters, whereas noise generally has higher spatial frequencies. Smoothing also facilitates between subject averaging as it blurs activations such that functionally homologous regions of individual subjects overlap. Finally, smoothing conditions the data to conform more closely to the Gaussian field.
Chapter 2 Methods

model, necessary for making statistical inferences about the regionally specific effects.

2.4.4 Statistical inference

In order to make statistical inferences about the data it has to be modeled to partition observed neurophysiological responses into components of interest (experimental conditions), confounds of no interest (e.g. global changes in cerebral blood flow) and an error term (unknown effects e.g. low frequency drifts or shifts in the fMRI signal over time). This partitioning is effected using the general linear model to estimate the components in terms of parameters associated with a design matrix. The general linear model is an equation that relates the observed to the expected response variables in terms of a linear combination of expected components and some residual error. The general linear model is also referred to as ‘analysis of covariance’ or ‘multiple regression analysis’. User specified parameters are used to create the design matrix, which contains one column for each hypothesised effect and one row for each image volume. The contribution of each effect to the observed physiological response is estimated. These estimated contributions, or parameter estimates, can range from simple factors such as the mean activity associated with a particular condition to more complicated factors such as an interaction term in a multifactorial experiment. Regionally specific effects are framed in terms of differences among these parameter estimates and are specified using linear compounds or contrasts. Different contrasts can be used to test a variety of effects, which generally fall
into one of the design categories mentioned earlier, namely, subtraction, factorial or parametric designs.

Regionally specific responses are characterised by computing a $t$ statistic for each contrast (or difference in parameter estimates) in parallel for each and every voxel, to form a SPM \{t\}. The SPM \{t\} is then converted to a Gaussian field or SPM \{Z\}. The SPM \{t\} is typically thresholded at $Z = 3.09 \ (p<0.001)$ uncorrected for multiple comparisons and resultant areas are characterised in terms of peak height and/or spatial extent. This procedure yields a $Z$ score for each activated area and a $p$ value that has been corrected for multiple comparisons using a Bonferroni like procedure. If an anatomically constrained hypothesis about the effects in a particular brain region was made \textit{a priori}, then the $Z$ value in that region in the SPM\{Z\} can be used to test the hypothesis by confining the estimation of $p$ values to small regions of interest. This is known as a small volume correction (Worsley et al., 1996). The convention adopted in the work presented in this thesis is to assess significance based on uncorrected $p$ values if the region of interest has been predicted \textit{a priori} before the data was collected. For all other areas and activations not predicted \textit{a priori}, corrected $p$ values are used.

The resulting statistical parameters are assembled into a statistical parametric map, in standard anatomical space (Talairach and Tournoux, 1988). SPMs are brain images with voxel values that are, under the null hypothesis, distributed
according to a known probability density function. ‘Unlikely’ excursions of the SPM are interpreted as regionally specific effects, attributable to the cognitive component that has been experimentally manipulated. SPM’s are images of the statistical reliability or significance of an effect, and these regionally specific effects or activations are shown as shaded voxels superimposed on the standard anatomical co-ordinate space. Activations are reported according to the location of the voxel of maximum intensity, which is given in Talairach and Tournoux (1988) co-ordinates, in which the origin (x,y,z = 0,0,0) is located at the anterior commissure. Co-ordinates are expressed in millimeters: x, distance to the right (+) or left (-) of the midsagittal plane; y, distance anterior (+) or posterior (-) to the vertical plane through the anterior commissure; z, distance above (+) or below (-) the intercommissural line.

2.5 Variation in sample size and gender distribution

The imaging studies in this thesis vary in sample size from 6 volunteers to 12 volunteers. The reason for this variation is due to the advances in the statistical methods occurring between the analysis of the first study in this thesis and the final imaging study. At the time of the first study (1998) it was recognised practice to use a fixed effects model using 6 volunteers however by 2001 it was possible to perform random effects analyses which require a minimum of 12 subjects. Also, chapters 3 and 6 used mixed gender groups, whereas only males were used in the study in chapter 4. Ideally, the use of both genders would be
preferable in studies of social cognition. However, the study in Chapter 4 used PET imaging which is regulated by the administration of radioactive substances advisory committee (ARSAC) which does not allow the participation of women under child bearing age.
Chapter 3: Reading the Mind in the Verbal and Nonverbal Domains.

3.1 Introduction

Chapter 1 discussed a number of theoretical viewpoints relating to the evolution and development of the cognitive mechanisms underpinning 'theory of mind', the human ability to attribute independent mental states to self and others in order to explain and predict behaviour. One theory, in particular, suggested that this ability arises from a dedicated, domain-specific, and possibly modular cognitive mechanism (Fodor, 1992; Leslie and Thaiss, 1992). This proposal gains particular support from studies of autism, a biologically based developmental disorder that appears to be characterised by a selective impairment in theory of mind (Happé and Frith, 1996). This chapter describes a functional neuroimaging study designed to test this theory by examining the neural correlates of 'theory of mind' ability in two modalities verbal and non-verbal (visual).

Prior to this study, there have been three published reports of functional brain imaging studies of 'theory of mind' on normal adults (Baron-Cohen et al., 1994; Goel et al., 1995; Fletcher et al., 1995a) and subsequently five others (Baron-Cohen et al., 1999; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001; McCabe et al., 2001). Baron-Cohen et al. (1994) used single photon emission computerised tomography (SPECT) and a regions of interest approach to isolate
brain areas activated during recognition of mental state terms in a word list. They found that their normal adult volunteers showed increased cerebral blood flow during the mental state recognition task in the right orbito-frontal cortex relative to the left frontal-polar region. Goel et al., (1995) used PET to scan adults engaged in a complex task in which volunteers had to model the knowledge and inference of another mind concerning the function of unfamiliar and familiar objects. They found widespread activation associated with this task, including activation of left medial frontal lobe and left temporal lobe. Fletcher et al. (1995a) also used PET, and scanned volunteers asked to read and answer questions about stories. Comparison of activation during "theory of mind" stories (requiring mental state attribution) versus control "physical" stores revealed task-specific activation in the left medial prefrontal gyrus, as well as increased activation of the posterior cingulate cortex. Using the same technique with individuals with a form of autism (Asperger Syndrome), Happé et al. (1996) found similar patterns of activation in all regions except the medial frontal gyrus which had been linked to theory of mind performance in the normal group.

Subsequent studies have also activated the medial prefrontal cortex in association with intention attribution or theory of mind (Baron-Cohen et al., 1999; Brunet et al., 2000; Castelli et al., 2000 Vogeley et al., 2001, McCabe et al., 2001, Chapter 5). Baron-Cohen et al. (1999) asked volunteers to make judgements about what a person might be thinking or feeling from the expressions of their eyes. They compared this with judgements about the gender of the persons depicted in the
stimuli. Brunet et al. (2000) studied intention attribution with a non-verbal version of the story comprehension task using comic strips. One condition required volunteers to attribute intentions to the characters of the comic strips. The other two conditions involved only physical logic and knowledge about the properties of objects, with and without characters. They found widespread activation when the attribution of intention condition was compared with the physical logic condition including the right medial prefrontal cortex (BA 8/9). Castelli et al. (2000) also found a network of regions associated with attributing intentions to the movements of geometric shapes based entirely on their kinematic properties. In two of the conditions the shapes provoked either mental state attribution, i.e. feeling states of the characters or goal directed behaviour. In both these conditions intentions were attributed to the shapes. In the third condition the motion of the shapes was random. Castelli et al. (2000) found medial prefrontal activity when both the mentalising condition and the goal directed movement condition were compared with the random movements. However this activation was more significant when associated with mentalising.

The main aims of the current study were to investigate the neural correlates underlying theory of mind by exploiting the superior spatial resolution of fMRI compared to PET and to examine anatomical convergence between theory of mind tasks presented in different modalities. The story comprehension task used by Fletcher et al. (1995a) was modified for compatibility with fMRI to examine theory of mind in the verbal domain, while captionless cartoons provided a visual
equivalent. Previous behavioural studies have used both stories (Frith and Corcoran, 1996; Happé, 1994; Happé et al., 1999) and visual jokes (Corcoran et al., 1997; Happé et al., 1999) to investigate theory of mind impairments in adults and children and found these tasks to be good markers of mentalising abilities. It was hypothesised that activation in medial prefrontal cortex would be associated with the attribution of mental states independent of modality. The final aim was to identify any modality specific regions for theory of mind in the verbal and visual domains.

3.2 Methods

3.2.1 Participants

Six right-handed volunteers with no neurological or psychiatric history participated in this study. Of these, five were male and one female, with a mean age of 30 years (range 23 to 36 years). This study was approved by the Institute of Neurology Ethics Committee. Informed written consent was obtained from all volunteers prior to scanning.

3.2.2 Tasks

All stimuli were displayed on a monitor and presented to the volunteer via a 45° angled mirror positioned above the head coil; this mirror was adjusted to be within the volunteer’s field of vision without having to tilt his/her head. A test
image was presented on the screen prior to scanning to ensure that the image was in focus and the volunteer could comfortably read the text.

3.2.2.1 Story comprehension task:

Three types of verbal material were presented; these were "theory of mind stories" (ToMS), "non-theory of mind stories" (Non-ToMS) and "unlinked sentences" (US). Figure 3-1. shows examples of these three types of material. During each scanning epoch a passage of text was presented on the screen for 21.6 seconds. The passage was then replaced by a question relating to the text (presented for a further 11 seconds). Volunteers were instructed to read the passage silently and to answer internally the question which followed. Each passage was preceded by a relevant prompt to indicate either "theory of mind", "unlinked sentences" or "non-theory of mind". This prompt was displayed for 1 second. The question was replaced by the next prompt indicating the start of a new epoch.

Following the scanning session each volunteer was shown the same passages again and asked to give his/her answer to each question, to provide a measure of behavioural performance. Responses were scored 1 for a correct answer and 0 for an incorrect answer. For the theory of mind condition an answer was considered correct only if an appropriate mental state was attributed to one or more characters. The maximum score was 4 per condition which was reached by all volunteers, thus indicating full understanding.
Chapter 3 Reading the mind in the verbal and nonverbal domains

Figure 3-1.

Figure 3-1-1. Theory of mind story.

A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn't know the man is a burglar, he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar, "Hey, you! Stop!", the burglar turns round, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.

Question: Why did the burglar do that?

Figure 3-1-2. Non-theory of mind story.

Two enemy powers have been at war for a very long time. Each army has won several battles, but now the outcome could go either way. The forces are equally matched. However, the Blue army is stronger than the Yellow army in foot soldiers and artillery. But the Yellow army is stronger than the Blue Army in air power. On the day of the final battle, which will decide the outcome of the war, there is heavy fog over the mountains where the fighting is about to occur. Low-lying clouds hang above the soldiers. By the end of the day the Blue army has won.

Question: Why did the Blue army win?

Figure 3-1-3. Unlinked sentences.

He needs a new engine for his old car. The prize is an immediate lump sum of £5,000 tax-free. Japan is stronger than Italy in economic terms. The mother is very brave and long suffering. The new book is about statistics and experimental design, and contains many graphs. The front room contained a little bird in a cage. Although Jim is only twenty one years old, he has an income of £20,000 per year. There are not many people this evening in the large rectangular dining room.

Question: Is the mother brave?
Chapter 3 Reading the mind in the verbal and nonverbal domains

The passages used in these tasks have been used in two previous experiments by this group (Fletcher et al., 1995a; Happé et al., 1996). The results from this study confirmed that the tasks were matched for difficulty, as reflected in scores and reading times, with ToM stories taking if anything slightly (but nonsignificantly) less time to read than the Non-ToMS.

3.2.2.2 Cartoon Task

Three types of picture were presented, analogous to the three types of text, "theory of mind cartoons" (ToMC), "non-theory of mind cartoons" (Non-ToMC) and "jumbled pictures" (JC). A cartoon was considered to require theory of mind for its interpretation if attribution of either false belief or ignorance to one or more of the characters in the picture was vital for comprehension. A cartoon was considered to be non-theory of mind if no mental state attribution was needed to understand the meaning. The "jumbled pictures" were constructed from images of randomly positioned objects, animals and people, taken from cartoons and children's colouring books. All images were captionless. Figure 3-2 shows examples of the three stimulus types.

The cartoons were validated in a pilot study with twenty naive normal volunteers (age range 19 - 66 years). Volunteers were instructed to look at each cartoon and indicate to the experimenter as soon as they understood its meaning. Response time was recorded. Volunteers then gave a brief explanation of the cartoon's meaning. The explanations were scored 1 for a correct answer and 0 for an
incorrect answer. As for the story condition, explanations in the theory of mind condition were considered correct only if an appropriate mental state was attributed to one or more characters. The explanation was recorded along with the time taken. In addition, participants were asked to rate, from 1 to 5, how difficult and how funny they thought the cartoon to be (1 extremely easy, 5 extremely
difficult; 1 meaning not funny and 5 extremely funny). No significant differences were seen in any of the measures between the two conditions.

During each scanning epoch four pictures were presented on the screen each for 8.15 seconds. Volunteers were instructed to look at each image during the theory of mind and non-theory of mind epochs and to consider the meaning silently. Volunteers were also asked "just to look" at the jumbled pictures in the control epochs. Each epoch was preceded by a relevant picture prompt, lasting 1 second, of either a brain, indicating a theory of mind epoch, a brain with a large cross through it, indicating a non-theory of mind epoch, or a face, indicating a jumbled picture epoch. The volunteers were shown the prompts prior to scanning to introduce them to this convention. The fourth image of each epoch was replaced by the next prompt indicating the start of a new epoch.

Following the scanning session each volunteer was shown the same cartoons again and asked to explain the meaning of each to provide a measure of performance. Volunteers scored 1 for a correct answer and 0 for an incorrect answer. As in piloting, attribution of mental states was required for explanation of ToM cartoons to be considered correct. The maximum score possible was 28 per condition. Once again the volunteers' performance was close to ceiling on both the ToMC and Non-ToMC conditions. Mean scores were: ToMC, 26.5 ($SD$, 0.84), Non-ToMC, 26.5 ($SD$, 2.14).
3.2.3 Data Acquisition

A Siemens VISION MRI system operating at 2 Tesla was used to acquire both T1 weighted anatomical and echo-planar T2* weighted image volumes with blood oxygenation level-dependent (BOLD) contrast. Functional images were acquired over two separate runs, a story run and a cartoon run, the order of the tasks and conditions was counterbalanced across volunteers. Each image volume constituted 48 3mm axial slices with in-plane resolution of 3mm by 3mm positioned to cover the whole brain. Volumes were acquired continuously every 4200ms while volunteers performed three experimental tasks, each task epoch comprised 8 image volumes. The story run comprised four theory of mind epochs and four non-theory of mind epochs, the cartoon run included seven theory of mind epochs and seven non-theory of mind epochs. Activation epochs were interspersed with control (rest) conditions. Each run began with six volumes which were discarded prior to analysis to allow for T1 saturation effects. A total of 364 volumes were acquired of which 352 were analysed. The duration of the scanning was approximately 40 to 45 minutes.

3.2.4 Data Analysis

Data were analysed using Statistical Parametric Mapping (SPM97, Wellcome Dept. of Cognitive Neurology, London, UK) implemented in MATLAB (Mathworks Inc., Sherborn, MA, USA) and run on a SPARC workstation (Sun Microsystems Inc., Surrey, UK). The imaging time series was realigned using the first image and spatially normalised to the stereotactic space of Talairach and
Toumoux (1988) using MNI templates (Montreal Neurological Institute). These data were subsequently smoothed with an isotropic Gaussian kernel of 9mm at full width half maximum.

Analysis was carried out using the general linear model and a delayed boxcar waveform. Subject-specific low-frequency drift in signal was removed by a high pass filter and global signal changes were removed by including a global covariate (Holmes et al., 1997). Effects at each voxel were estimated and regionally specific effects were compared using linear contrasts. The resulting set of voxel values for each contrast constituted a statistical parametric map of the t statistic (SPM(t)) which was subsequently transformed to the unit normal distribution, SPM(Z). Statistical inferences were based on the theory of random Gaussian fields (Friston et al., 1995). Activations are reported significant at \( p < 0.05 \) corrected for multiple comparisons. In regions about which there is an a priori hypothesis, activations are considered significant at \( p < 0.001 \) uncorrected (Friston et al., 1995).

The stereotactic coordinates of Talairach and Tournoux, 1988) are used to report the observed activation foci. However, descriptions of the anatomical localisation of the foci were determined using averaged structural MRIs of the group and the atlas of Duvernoy (1991).
3.3 Results

3.3.1 Theory of Mind Stories versus Non-Theory of Mind Stories.
Significant activations were seen in the medial frontal cortex, the temporal poles bilaterally and the temporo-parietal junctions bilaterally (exact coordinates are given in table 3-1).

3.3.2 Theory of Mind Stories versus Unlinked Sentences.
Significant activations were seen in the medial frontal cortex, the temporal poles bilaterally and the temporo-parietal junctions bilaterally.

3.3.3 Non-Theory of Mind Stories versus Theory of Mind Stories.
No significant activations were seen in this contrast.

3.3.4 Non-Theory of Mind Stories versus Unlinked Sentences.
Significant activations were seen in the temporal poles bilaterally and the temporo-parietal junctions bilaterally.

3.3.5 Theory of Mind Cartoons versus Non-Theory of Mind Cartoons.
Significant activations were seen in the medial prefrontal cortex, the temporo-parietal junctions bilaterally, the right middle frontal gyrus, the precuneus and the cerebellum (table 1).
3.3.6 Theory of Mind Cartoons versus Jumbled Pictures.

Significant activations were seen in the medial prefrontal cortex, the temporoparietal junctions bilaterally, the right middle frontal gyrus, the precuneus and the cerebellum.

3.3.7 Non-Theory of Mind Cartoons versus Theory of Mind Cartoons.

No significant activations were seen in this contrast.

3.3.8 Non-Theory of Mind Cartoons versus Jumbled Pictures.

Significant activations were seen in the temporal poles bilaterally and the temporoparietal junctions bilaterally.


Significant activations were seen in the medial prefrontal cortex and the temporoparietal junctions bilaterally, in the comparison of ToM and Non-ToM stimuli combining both the story and cartoon tasks (table 3-1) (fig 3-3). Other regions that activated but did not reach a corrected level of significance were the right middle frontal gyrus, the precuneus, and the left temporal pole.
Table 3-1
Regions of increased brain activity associated with theory of mind compared with non-theory of mind for i. story comprehension, ii. cartoons, and iii. story comprehension and cartoons.

<table>
<thead>
<tr>
<th>Region</th>
<th>Putative Brodmann area</th>
<th>*x</th>
<th>y</th>
<th>z</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ToM vs non-ToM stories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal gyrus</td>
<td>8/9/32</td>
<td>-8</td>
<td>50</td>
<td>10</td>
<td>3.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>48</td>
<td>30</td>
<td>3.49</td>
</tr>
<tr>
<td>L temporal pole</td>
<td>38</td>
<td>-48</td>
<td>14</td>
<td>-36</td>
<td>4.15</td>
</tr>
<tr>
<td>R temporal pole</td>
<td>38</td>
<td>54</td>
<td>12</td>
<td>-44</td>
<td>3.66</td>
</tr>
<tr>
<td>L temporo-parietal junction</td>
<td>39/40</td>
<td>-46</td>
<td>-56</td>
<td>26</td>
<td>4.04</td>
</tr>
<tr>
<td>R temporo-parietal junction</td>
<td>39/40</td>
<td>66</td>
<td>-52</td>
<td>8</td>
<td>3.65</td>
</tr>
<tr>
<td><strong>ToM vs non-ToM cartoons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal gyrus</td>
<td>8/32</td>
<td>4</td>
<td>26</td>
<td>46</td>
<td>3.1</td>
</tr>
<tr>
<td>R middle frontal gyrus</td>
<td>6</td>
<td>40</td>
<td>8</td>
<td>42</td>
<td>5.16</td>
</tr>
<tr>
<td>R temporo-parietal junction</td>
<td>40</td>
<td>58</td>
<td>-44</td>
<td>24</td>
<td>5.39</td>
</tr>
<tr>
<td>Precuneus</td>
<td>7/31</td>
<td>12</td>
<td>-52</td>
<td>58</td>
<td>4.49</td>
</tr>
<tr>
<td>Fusiform</td>
<td>20/36</td>
<td>46</td>
<td>-44</td>
<td>-24</td>
<td>4.18</td>
</tr>
<tr>
<td><strong>ToM vs non-ToM stories &amp; cartoons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal gyrus</td>
<td>8/9/32</td>
<td>-10</td>
<td>48</td>
<td>12</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>22</td>
<td>46</td>
<td>3.15</td>
</tr>
<tr>
<td>R middle frontal gyrus</td>
<td>6</td>
<td>42</td>
<td>8</td>
<td>46</td>
<td>4.02</td>
</tr>
<tr>
<td>L temporal pole</td>
<td>38</td>
<td>-48</td>
<td>16</td>
<td>-38</td>
<td>4.01</td>
</tr>
<tr>
<td>L temporo-parietal junction</td>
<td>39/40</td>
<td>-54</td>
<td>-66</td>
<td>22</td>
<td>5.18</td>
</tr>
<tr>
<td>R temporo-parietal junction</td>
<td>39/40</td>
<td>60</td>
<td>-46</td>
<td>22</td>
<td>5.42</td>
</tr>
<tr>
<td>Precuneus</td>
<td>7/31</td>
<td>2</td>
<td>-50</td>
<td>44</td>
<td>3.45</td>
</tr>
</tbody>
</table>

*Coordinates are given with reference to a standard stereotactic space (Talairach and Tournoux, 1988).

3.310 Interaction: (Theory of Mind versus Non-Theory of Mind) x (story comprehension versus cartoons).

This interaction demonstrated regions of increased activity associated with theory of mind, specific to the story comprehension task. Only one region, the medial prefrontal cortex was shown to be significantly activated (table 3-2).
3.3.11 Interaction: (Theory of Mind versus Non-Theory of Mind) x (cartoons versus story comprehension).

This interaction demonstrated regions of increased activity associated with theory of mind specific to the cartoon task. Significant activations were seen in the right middle frontal gyrus, the precuneus, and the cerebellum (table 3-2).

Table 3-2
Regions of increased brain activity associated with theory of mind compared with non-theory of mind specific to i. story comprehension and ii. cartoons.

<table>
<thead>
<tr>
<th>Region</th>
<th>Putative Brodmann area</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. Interaction: (ToM vs non-ToM) x (stories vs pictures)</td>
<td>Medial prefrontal gyrus</td>
<td>9/32</td>
<td>10</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>ii. Interaction (ToM vs non-ToM) x (pictures vs stories)</td>
<td>R middle frontal gyrus</td>
<td>6</td>
<td>54</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td>6</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precuneus</td>
<td>7/31</td>
<td>2</td>
<td>-56</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Flocculus</td>
<td>-24</td>
<td>-30</td>
<td>-42</td>
<td></td>
</tr>
</tbody>
</table>


Figure 3-3. A statistical parametric map (SPM{Z}) of the group data as a maximum intensity projection showing the areas where there was greater activation to theory of mind stories and cartoons compared to non-theory of mind stories and cartoons. Views are from the right, top and behind.
3.4 Discussion

This study sought to clarify further the functional anatomy of "theory of mind" using fMRI. It attempted to examine anatomical convergence between "theory of mind" tasks in two domains, verbal and visual, and identify modality specific regions for "theory of mind" in these domains. The results of this experiment corroborate the evidence of previous studies of story comprehension, either verbal (Fletcher et al., 1995a; Vogeley et al., 2001) or non-verbal (Brunet et al., 2000). They are also consistent with the results of two different studies by Goel et al. (1995) and Baron-Cohen et al. (1999), suggesting that the ability to mentalise is mediated by the medial prefrontal cortex. In addition, these results suggest that this region is activated by ToM tasks regardless of modality. A conjunction analysis of ToM versus Non-ToM activation common to both tasks, also demonstrated increased activity of the temporo-parietal junctions bilaterally. However, the medial prefrontal cortex was the only region uniquely activated in the theory of mind condition. All other regions were also activated (albeit to a significantly lesser extent) in the Non-ToM versus control contrast.

3.4.1 Medial prefrontal cortex (BA 8/9/32)

The critical area activated in association with theory of mind was medial prefrontal cortex. By examining each individual subject’s activations plotted onto his/her own T1 weighted structural images this region was pinpointed to the
medial convexity labelled BA 8/9 by Talairach & Tournoux (1988), but closely associated with the anterior cingulate region of BA 32 (Fig 3-4). This is consistent with previous PET studies of theory of mind that have referred to the region as BA 8/9, predominantly on the left (Fletcher et al., 1995a; Goel et al., 1995). Since the cytoarchitecture of this region is not well defined, the area of activation would be more appropriately labelled as paracingulate cortex (R. Passingham, personal communication). In those volunteers with two or more cingulate sulci, the activity was seen in the paracingulate sulcus. A similar region was shown by Bottini et al. (1994) to be activated when processing of metaphorical sentences was compared with processing of literal sentences.

The verbal ToM task activated a broader region of the medial prefrontal cortex which extended anteriorly and inferiorly. According to the atlas of Talairach & Tournoux (1988) activation elicited by ToM cartoons was restricted to Brodmann’s area 8/32, while that associated with ToM stories extended into the adjacent area 9/32. The cartoon and story tasks in the present study were not equated for difficulty, and may have differed in the level of theory of mind or degree of mental state embedding which they elicited from participants. Whilst involvement of the medial prefrontal cortex was demonstrated in both, the extent of the activation was greater in the verbal task. This may have reflected possible differences in the subtracted Non-ToM task; in viewing cartoons, participants may try to work out what the cartoonist intended the joke to mean - and this may lead to a degree of theory of mind activity even during viewing of cartoons.
without mental state content. The cartoon task was associated with increased activity in additional regions; the precuneus (BA7), the middle frontal gyrus (BA6) and the cerebellum. Once again, these regions activated in Non-ToMC compared to the control condition (JC) and therefore cannot be regarded as specific to theory of mind.

Figure 3-4 Area of activation in the medial frontal cortex of a single volunteer elicited by theory of mind stories and cartoons. Co-registration of functional and structural scans for this volunteer show that the activation lies in the paracingulate cortex.
3.4.2 Temporo-parietal junctions (BA 39/40)

Widespread increased activation of the temporo-parietal junctions bilaterally (but more significantly on the right) was found to be common to both the story comprehension and the cartoon tasks. Three subsequent studies of theory of mind have also found activity in this region (Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001). Brunet et al., 2000 found right superior temporal sulcus (STS) to be associated with the involvement of people compared to physical objects alone and Castelli et al. (2000) also found right STS associated with attributing intentions, but not necessarily mental states, suggesting that this region is not involved in mentalising per se. Left sided activations of this region are often seen in imaging studies of language processing and have been attributed to semantic knowledge of single words, (x, y, and z, coordinates: -40, -70, 24) (Price et al., 1997; Vandenbergh et al., 1996; Warburton et al., 1996). This is borne out by lesion data (Alexander et al., 1989; Damasio et al., 1996). Bilateral activations of a portion of this same region were elicited by Puce et al. (1998) (x, y, and z, coordinates: 51, -49, 5) when volunteers perceived movements of the eyes and mouth compared to non-facial movement in the same part of the visual field or movement of a radial background. An area posterior to that of Puce’s (1998) but within the boundaries of the region of the current study, was also activated by Bonda et al. (1996) (x, y, and z, coordinates: 56, -54, 8) by the perception of simulated hand actions and body movements compared to object and random motion. These two studies have been interpreted as showing a role for this region in the perception of biological motion. The theory of mind
materials in this study, in which there was no motion, also activated these regions. This may suggest that the region of the temporo-parietal junction is sensitive not merely to biological motion but, more generally, to stimuli which signal intentions or intentional activity. Frith and Frith (1999) have speculated that this region may be concerned more generally with the recognition of intentions. They have proposed that the mechanism allowing the development of a theory of mind is derived from the interaction of the ability to monitor one’s own mind (paracingulate cortex) and the ability to monitor the behaviour of other living things (R. temporo-parietal junction) which they suggest is an adaptation of a more primitive social brain. However, Vogeley et al., (2001) used a modification of the story comprehension task from this study to add a new dimension to the test. They looked specifically for the existence of differences between monitoring one’s own mind (self-perspective) and monitoring the minds of others (mindreading). They found the right temporo-parietal junction was specifically involved in taking the self-perspective. They suggest that in the current study the stimuli were not controlled for volunteers implicitly taking the self perspective.

3.4.3 The right middle frontal gyrus (BA 6)

In the current study a region of the right middle frontal gyrus (BA 6) was also activated during the consideration of cartoon meaning, and particularly theory of mind cartoon meaning. This corresponds closely to another region activated by Bottini et al. (1994) in association with processing metaphorical sentences compared with processing literal sentences. This suggests a role for the attribution
of mental states during interpretation of metaphorical utterances. Indeed, Happé (1993) has demonstrated a strong theoretical and empirical relation between theory of mind capacities and understanding of figurative language (metaphor, sarcasm) in normal children and those with autism. A role for right hemisphere regions in understanding of mental states is supported by neurological studies showing impairments on theory of mind tasks following right hemisphere stroke (Siegal et al., 1996; Happé et al., 1999).

3.4.4 The medial parietal (precuneus) cortex (BA 7/31)

The medial parietal (precuneus) region was activated in all the conditions when compared with the appropriate baseline control. However, this region was shown to be significantly more activated in the ToMC than the Non-ToMC task. Previous functional imaging studies have suggested a role for the precuneus in mental imagery at the retrieval stage of episodic memory (Shallice et al., 1994; Fletcher et al., 1995b). The study by Fletcher et al. (1995a), which used only the story comprehension task, reported a deactivation of the precuneus in the comparison of unlinked sentences versus stories. Lowering the threshold in this current study demonstrated activity associated with mentalising in the story comprehension task. This may indicate that the deactivation in the original paper resulted from a subthreshold activation in the ToMS versus unlinked sentences comparison, and suggests increased use of mental imagery in theory of mind tasks.
3.4.5 The left flocculus of the cerebellum and the fusiform.

The interaction condition showed activations in the left flocculus of the cerebellum. There is a growing body of literature suggesting a role for the cerebellum in higher cognitive functions such as non-motor learning, executive function, spatial cognition, language and emotional regulation of behaviour (Raichle et al., 1994; Leiner et al., 1995; Schmahmann and Sherman, 1998). However, in this instance it may result from the use of differential strategies for scanning pictures and text, which is consistent with previous studies associating the flocculus with the control of eye movements (Krauzlis and Lisberger, 1996). Finally, increased activity of the fusiform was shown to be associated with understanding the meaning of visual jokes in this study, particularly during the theory of mind task. This region is known to be associated with the processing of faces and objects (Bly and Kosslyn, 1997).

3.5 Conclusions

This chapter has reported a study of the functional anatomy of theory of mind, using tasks in visual and verbal modalities. Story and cartoon tasks requiring mental state attribution engaged specific networks of cortical regions, and showed common areas of increased activation in the medial prefrontal gyrus and the temporo-parietal junctions bilaterally. An area of medial prefrontal cortex (the anterior paracingulate cortex) was the only region uniquely activated by theory of mind tasks, and not activated above baseline in the comparison stories or
cartoons. This provides further evidence that the ability to attribute mental states is mediated by this highly circumscribed brain region, and that such activation is independent of modality.

One potential criticism of this study (and earlier functional imaging studies of theory of mind) is that it uses an off-line task that may not involve the same mechanisms as those occurring in real-time. In this study the volunteers are retrospectively explaining the behaviour of the characters in the stories and cartoons. Mentalising is a highly automatic process. It allows us to both explain and predict the behaviour of others. Chapter 4 discusses a study of real-time theory of mind using an on-line mentalising task to address this issue.
4 CHAPTER 4: IMAGING THE INTENTIONAL STANCE

4.1 Introduction

The previous chapter described an fMRI study that examined the neural correlates of theory of mind ability in two modalities, verbal and non-verbal. Theory of mind is our ability to explain and predict the behaviour of others by attributing independent mental states, such as thoughts, beliefs, desires and intentions, different from our own. The results of the study described in chapter 3 showed that a small number of brain areas, including the anterior paracingulate cortex, are activated by mentalising tasks regardless of modality. This is consistent with previous literature of functional imaging studies of normal volunteers (Baron-Cohen et al., 1994; Goel et al., 1995; Fletcher et al., 1995a; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001, McCabe et al., 2001) which have also sought to examine the neural substrates of mentalising ability. However, mentalising is a complex higher cognitive function, and a ubiquitous element of social reasoning and problem solving. Thus, it is a process that is not easily isolated. The paradigms employed in previous imaging studies of mentalising (Baron-Cohen et al., 1994; Goel et al., 1995; Fletcher et al., 1995a; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001 McCabe et al., 2001), have used a range of tasks and conditions involving additional cognitive demands. Not surprisingly, these studies have implicated multiple neural regions, but there is also a high degree of overlap.
between studies. In the majority (Goel et al., 1995; Fletcher et al., 1995a; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000, Vogeley et al., 2001) three brain regions have been consistently activated which appear to form the basis of a neural network of social perception (Frith, 2001). These regions are 1) the anterior paracingulate cortex (also activated in McCabe et al., 2001), 2) the posterior part of the superior temporal sulcus at the temporo-parietal junction, and 3) the temporal pole. However, the precise roles of these regions, with regard to the specific cognitive mechanisms involved in these tasks of mentalising, remain unclear. Furthermore, all of the previous studies, with the exception of McCabe et al., 2001, have employed “off-line” tasks which require the volunteer to consider a scenario and retrospectively explain the behaviour of the person or persons involved. This raises the question: Are the same mechanisms and regions of the brain involved in on-line, or in other words real life, mentalising situations? This chapter addresses this question with an examination of real-time theory of mind using an “on-line” mentalising task.

Thus the principal aim of the current experiment was to see whether the same brain regions were activated during “on-line” mentalising. The mentalising task used in this study was the childrens’ game ‘Stone-paper-scissors’. Playing this game required volunteers to predict the actions of their opponent in real time. This task was compared with two control tasks that did not require mentalising. In one of these tasks volunteers played the same game against a computer which used
simple rules and in the other they generated a random sequence of moves (see
methods and Table 4-1). If ‘on-line’ mentalising (actively predicting another
agent’s actions) involves the same processes as ‘off-line’ mentalising
(retrospectively explaining another agent’s actions) then the critical comparison
should activate regions of overlap with previous studies.

The second aim of the current study was to isolate a particular component of
mentalising; the adoption of an intentional stance (Dennett, 1971; Dennett, 1991;
Dennett, 1996). That is, volunteers would conceive of their opponent as a rational
agent, possessing particular beliefs, desires and intentions, and this understanding
would form the basis for their own actions in the game.

During acquisition of all functional data, volunteers were, without their
knowledge, playing against a random sequence. The only actual differences
between conditions were due to differences in the mental attitude or ‘stance’ of
the volunteers. The shift between playing against a computer and playing against
a person, and therefore adopting an ‘intentional stance’ was induced using explicit
verbal instructions and through context. Thus, lead in and lead out trials, in which
the volunteers were actually playing the experimenter or a computer, flanked the
random sequence, lulling them into believing the manipulation. The aim of this
manipulation was to influence the tendency for subjects to mentalise, whilst
minimising other cognitive differences. In previous studies of mentalising, volunteers not only had to adopt an intentional stance, but also had to process different kinds of signals. In the current study the signals to be processed did not differ, only the attitude that the volunteers took towards these signals. Regions activated in common with these previous studies should therefore be specifically concerned with the stance taken rather than the kind of signals being processed.

However, there is clearly a danger that volunteers may not have responded in the manner intended. In previous imaging studies, the only assurances of the success of this type of experimental manipulation have been implicit and/or post-hoc. For example, a post-hoc finding that activation is specific to a single region might be read as supporting the starting assumption, that the manipulation specifically affects the cognitive process under investigation. However, 'reverse inferences' of this sort are question begging and prone to error (D'Esposito et al., 1998). Therefore, it was decided that some independent evidence should be sought concerning the effects of the experimental manipulation. This was achieved by collecting volunteers' own accounts of how they understood and experienced the different conditions. This approach was perceived as having two advantages: First, it would provide information about the demands of the different conditions, and any differences in the strategies used by volunteers. Second, it would provide information on ‘what it is like’ to adopt intentional and physical stances. Thus the aim was to identify the phenomenology associated with the experiment, in order
to determine whether there was a general experience that could be correlated with
neuronal activation.

4.2 Materials and Methods

4.2.1 Subjects

Nine healthy male volunteers, mean age of 32 years (range 22 to 43 years)
participated in this study. They all gave informed consent and the study was
approved by the Ethics Committee of the Institute of Neurology, UCL and
ARSAC (UK). Informed written consent was obtained from all volunteers prior to
scanning.

4.2.2 Data Acquisition

Each volunteer underwent 12 PET relative perfusion scans over a 2-hour period.
Scans were obtained using a Siemens/CPS/ECAT EXACT HR+ (model 962) PET
scanner (Siemens/CTI, Knoxville, Tenn., USA) with collimating septa retracted.
Participants received a 20-s intravenous bolus of $H_2^{15}O$ at a concentration of 55
MBq/ml and at a flow rate of 10ml/min through a forearm cannula. Functional
images were acquired over a 90-s period. All measurements were corrected for
attenuation by transmission scans of 5-min duration. In addition, $T_1$-weighted
volume acquired images was obtained for each volunteer with a Siemens
Chapter 4  
Imaging the intentional stance

Magnetom VISION MRI system operating at 2 Tesla (Siemens, Erlangen, Germany).

4.2.3 Data Analysis

PET data were analysed using Statistical Parametric Mapping (SPM99d) software from the Wellcome Dept. of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB (Mathworks Inc., Sherborn, MA, USA) and run on a SPARC workstation (Sun Microsystems Inc., Surrey, UK) using standardised procedures (Friston et al., 1995a). The images were smoothed using an isotropic Gaussian kernel (FWHM 8mm). Condition and subject effects were estimated according to the general linear model at each voxel. To test hypotheses about regionally specific condition effects, the estimated effects were compared using linear compounds or contrasts. The resulting set of voxel values for each contrast constitutes a statistical parametric map of the $t$-statistic, $SPM(t)$. The $SPM(t)$ were transformed to the unit normal distribution $SPM(Z)$. Activations are reported significant at $p < 0.05$ corrected for multiple comparisons. In regions about which there is an apriori hypothesis, activations are considered significant at $p < 0.001$ uncorrected. The stereotactic coordinates of Talairach and Tournoux (1998) are used to report the observed activation foci. However, descriptions of the anatomical localisation of the foci were determined using averaged structural MRIs of the group and the atlas of Duvernoy (1999).
4.2.4 Experimental Paradigm

Stimuli were displayed on a monitor and presented comfortably within the volunteer’s field of vision. The task consisted of a computer-based version of the playground game “Stone, paper, scissors!” Stone beats scissors, paper beats stone and scissors beats paper. In this version the selection was made on a keyboard and the choice displayed as an icon in the result window (see Fig 4-1). The result window showed the volunteer’s choice on the left above the word “YOU” and his opponent’s on the right above the word “PLAYER”. A star displayed behind the winning choice. In the event of a draw a diamond was displayed behind each choice to maintain the visual input.

Figure 4-1. Diagrammatic representation of the display and timing of the stimuli. For each trial the screen counted down 3, 2, 1 at 0.5 second intervals “GO” was displayed for a further second then replaced by a blank screen for a 0.5 seconds. The result window was displayed for 4 seconds. One trial lasted 7 seconds. The volunteers made their selection while GO was displayed on the screen.
If the volunteer failed to make a selection, their screen remained blank and the opponent was credited with a win. Each volunteer played a total of 12 “games”, 1/scan run - four games of each condition. Each “game” consisted of 30 trials. Instructions were given prior to scanning and training games were played to familiarise the volunteers with the display and timing (Fig 4-1), the response keyboard and the rules of each condition.

Since some people believe that the most effective method of play is to be as random as possible volunteers were explicitly instructed to try and out wit and second guess their opponent, when the condition was playing against a person.

In the other two conditions volunteers played against a computer. In the rule solving condition the computer based it’s response on a simple rule that was related to the volunteer’s previous responses. Thus the computer’s selections were a) the same as the volunteer’s last response, b) the response which would have beaten the volunteer’s last response or c) the response which would have lost to the volunteer’s last response. The rules were then to always use one of these selections (e.g. a a a a a etc.) or to alternate between two of them (e.g. a b a b a b etc.). The rule that the computer used changed a few times during the run. Volunteers were warned of this and asked to readjust to the new rule when they realised it had changed. Volunteers were instructed that the aim of the mentalising
and rule solving conditions was to win as many trials as possible, thus the volunteer’s score was calculated and displayed at the end of each run.

During the random selection condition the computer randomly generated its choice. The volunteer was asked to respond randomly but not to choose the same response as the computer had on the previous trial. The stars and diamonds indicating win, lose or draw were still displayed to maintain visual input but the volunteer was not given a score. This condition acted as a baseline task.

| Table 4-1. Higher cognitive processes thought to be engaged in each condition. |
|--------------------------------------|---------------------------------|----------------------------------|
| **Mentalising** | **Rule Solving** | **Random Selection** |
| Working Memory | Working Memory | Working Memory |
| Response selection | Response selection | Response selection |
| Random generation | Random generation | Random generation |
| Pattern Abstraction | Pattern Abstraction | |
| Strategy | Strategy | |
| Mentalising | | |

To isolate activity associated with the intentional stance a random sequence was inserted into each mentalising and rule solving game, which coincided with and lasted the duration of the critical scan window. Thus the volunteers played 15
lead-in trials against their anticipated opponent followed by 10 randomly
generated trials and 5 lead-out trials against the opponent. The purpose of the
lead-in and lead-out trials was to convince the volunteers of the presence of a
particular opponent. A continuous random sequence throughout the game may
have been detected. Volunteers were not informed of the random sequence.
Before commencement of each set of trials the volunteer was informed of the
condition and reminded of the appropriate rules. It was stressed to the volunteers
that they should not stereotype their responses in any of the conditions.

4.2.5 Behavioural Data

Verbal report data were collected in two phases. First, four pilot subjects were
given unstructured interviews during a development phase. These data were used
to refine the experimental paradigm and instructions, and to identify key questions
for interview during the scanning phase. Immediately after the experiment, the
volunteers were interviewied for 10-20 minutes. The interviews were conducted
semi-structurally, following a common list of questions, yet leaving the volunteers
room to find their own wordings and leaving the interviewer room to follow up on
particular aspects. The interviews were taped and subsequently transferred to
computer. The interviews were coded and annotated by two researchers using
HyperResearch 2.03 (ResearchWare, MA). The coders had no knowledge of the
scanning data. Subsequently, central parts of the interviews were transcribed for
further analysis.
4.3 **Results**

4.3.1 *Verbal reports*

The principle observations from the verbal reports collected after scanning were as follows:

1. No volunteer realized that a random sequence was inserted during the scanning window, in the computer or mentalising conditions.

2. In discussing the mentalising condition, volunteers unanimously depicted the opponent as an intentional agent. They all described guessing and second guessing their opponent's responses, and attempting to identify characteristic patterns of behaviour in the opponent. To a lesser extent, they described their opponent as reading their own patterns of behaviour. Volunteers spoke of strategies for response, which they related to imagined mental states of their opponent. However, there was one exceptional volunteer who claimed that the “nature of the game” was such that the notion of strategy did not make much sense since any strategy would immediately be detected by the opponent. Nonetheless, this volunteer did report that he performed worse against the person than against the computer and that he felt particularly disappointed when he lost against the person. Thus, in spite of the apparent similarity in strategy,
he experienced a difference when playing against a person rather than a computer.

3. There was considerable variation in volunteers’ descriptions of the computer condition. The instructions stated that the computer would follow rules that related its current response to the previous response of the volunteer. Some volunteers attempted to identify these simple rules, and occasionally succeeded. They described the computer as in principle very predictable, although the rules it used could not always be detected. Others found the computer’s responses more baffling, and contrasted it with playing the experimenter - whose actions they felt they could understand and ‘go along with’. These volunteers elaborated their own, generally vague, conceptions of what the computer was doing. One volunteer imagined that the computer was engaged in extremely complex multi-layered analyses of his patterns of response.

4. A number of volunteers used mentalistic language when depicting the computer. However, use of mentalistic language is not necessarily a good indicator of adopting an intentional stance. The clearest difference between physical and intentional stances is the subjective sense of another’s presence. When volunteers were asked to describe differences between computer and mentalising conditions, two consistent findings emerged. Firstly, all volunteers reported that the two conditions ‘felt’ distinctly different, and that the mentalising condition involved a feeling
Chapter 4

Imaging the intentional stance

doing interacting with another person whose actions one could 'go along with'.
Secondly, they found the computer condition more time pressured and strenuous than the mentalising condition. Volunteers explained this by reporting that they felt they could not keep up with a computer, but that when they were playing another person the timing was not so critical, since their opponent would have the same amount of time as them.

Verbal reports of this sort are not appropriate for quantitative analysis. However, they may assist in the interpretation of the main findings. Two points are particularly relevant. First, the reports testify that the conditions were experienced as quite different, although the actual stimuli in the scanning window were the same. The experience of playing the person was described as an interaction with a rational agent with distinct beliefs, intentions and desires. This contrasted with the computer, which was not given these attributions but was described as acting according to a design. The reports thereby support the claim that volunteers did indeed take 'an intentional stance' when playing against a person. Secondly, volunteers clearly experienced the computer condition to be more demanding than the mentalising condition. This finding is significant, given the suggestion that paracingulate activation may be associated with a subjective sense of effort (Bush et al., 2000).
4.3.2  Performance measures

Performance measures showed that there was no significant difference ($t(8) = 0.6$, $p = 0.5$) between the conditions during the critical scan window, (% mean wins; mentalising = 33.1, rule solving = 31.2); % mean losses; mentalising = 32.2, rule solving = 34.7; % mean draws; mentalising = 34.7, rule solving = 33.9) or during the pre-scan trials ($t(8) = 1.9$, $p = 0.08$) (% mean wins; mentalising = 24.4, rule solving = 37.3; % mean losses; mentalising = 30.9, rule solving = 34.3; % mean draws; mentalising = 44.7, rule solving = 28.4). There was also no significant difference between these conditions in the number of failures to respond during the critical scan window ($t(8) = 0.4$, $p = 0.7$; mean; mentalising = 1, rule solving = 1.22) or during the pre-scan trials ($t(8) = 1.069$, $p = 0.3$; mean; mentalising = 1.22, rule solving = 1.88). These results confirm that the only difference between the conditions lay in the stance of the volunteer.

4.3.3  Brain activity

The critical comparison, looking at where brain activity was greater in the mentalising condition than the computer condition, revealed only one region of significant difference; the most anterior portion of the paracingulate cortex bilaterally; Brodmann area 32/9 (see Table 4-2 & Fig. 4-2). No further regions appeared when the statistical threshold was lowered to $p=0.1$. The same region activated significantly when mentalising was compared with random generation.
In addition, this comparison revealed activity in the right inferior frontal cortex (BA 47) and the cerebellum (Table 4-2). The right inferior frontal cortex was also seen to be significantly more active in the computer condition as compared with random generation. No activity was seen in the paracingulate cortex in this comparison even when the threshold was lowered to p=0.1. As was expected no regions were significantly more active in the computer condition than in the mentalising condition.

<table>
<thead>
<tr>
<th>Table 4-2. Regions of increased brain activity associated with mentalising compared with (i) rule solving and (ii) random generation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Putative Brodmann Area</strong></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>(i) Mentalising vs rule solving</strong></td>
</tr>
<tr>
<td>Anterior paracingulate cortex</td>
</tr>
<tr>
<td>Anterior paracingulate cortex</td>
</tr>
<tr>
<td><strong>(ii) Mentalising vs random generation</strong></td>
</tr>
<tr>
<td>Anterior paracingulate cortex</td>
</tr>
<tr>
<td>Left inferior frontal cortex</td>
</tr>
<tr>
<td>Cerebellum</td>
</tr>
<tr>
<td>Cerebellum</td>
</tr>
</tbody>
</table>
Figure 4-2.
An SPM showing the group activation in the most anterior portion of the paracingulate cortex elicited in the contrast of mentalising versus rule solving in sagittal, coronal and axial views, mapped onto a template brain.

4.4 Discussion

4.4.1 Intentional stance

The main comparison, of the mentalising condition with the computer condition, revealed activity in just one of the three regions consistently observed in previous
functional imaging studies of mentalising in normal volunteers (Goel et al., 1995; Fletcher et al., 1995a; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000, Vogeley et al., 2001); the anterior paracingulate cortex. The tasks used in these previous studies were associated with many aspects of mental state attribution. For example, they involved desires, intentions, thoughts and knowledge (Fletcher et al., 1995a; Chapter 3, Brunet et al., 2000). On the other hand, in these tasks the mentalising was done ‘off-line’ and did not involve direct interaction with another person. Furthermore these tasks did not involve competition or winning and losing. The common feature of the tasks used in the present and these previous studies is the requirement to take an intentional stance; in other words to use the available information to discover the beliefs, intentions and desires of another person. Of particular relevance is the recent study by McCabe et al. (2001). As in the present study, McCabe et al. (2001) used an ‘on-line’ task that involved anticipating the behaviour of another player. However, unlike the present study the task was co-operative rather than competitive. In the critical comparison McCabe et al. (2001) also observed activity in anterior paracingulate cortex at co-ordinates very similar to those observed in the present study (x = 5, y = 52, z = 10 from Figure 4 in McCabe et al., 2001).

The absence of activity in the right temporo-parietal cortex in the present study is interesting as this region has been prominent in most previous imaging studies of mental state attribution (Goel et al., 1995; Fletcher et al., 1995a; Baron-Cohen et
al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000). However, this region is not necessarily associated with mentalising *per se*. For instance, Brunet et al. (2000) found bilateral temporal activation to be associated with the appearance of people in the cartoon strips even when it was not necessary to consider their intentions. The study in Chapter 3 used a story comprehension task and humorous cartoons to compare mentalising in both the verbal and the non-verbal domains. Bilateral temporo-parietal activation (particularly on the right) was found when the mentalising conditions were compared with equivalent non-mentalising conditions and a low-level control condition. However, activation in this region was also found, although less strong, when the non-mentalising conditions were compared with the low-level control, thus indicating that these are not uniquely associated with mentalising. The R temporo-parietal cortex has also been activated in previous imaging studies involving the perception of biological motion (Bonda et al., 1996; Calvert et al., 1997; Puce et al., 1998; Grezes et al., 1998; Grezes et al., 1999; Grossman et al., 2000; Grossman et al., 2001). A common factor of all those studies is that they all involve some explicit representation of behaviour, either directly or by verbal description. The behaviour of others is the principal source of information for making attributions about their mental states, but such information is only used in this way when we adopt an intentional stance. It seems plausible that this region of temporo-parietal cortex is activated by the presentation of signals important for mentalising even when the subject is not adopting an intentional stance. In contrast, anterior paracingulate cortex is activated when the subject adopts an intentional stance.
even when there are no additional verbal or visual cues to assist mentalising. Presumably the region of the temporal pole also activated in many studies of mentalising has a similar role to temporo-parietal junction; it is concerned with signals important for mentalising, but not with mentalising *per se*.

### 4.4.2 The role of anterior paracingulate cortex

There is evidence from studies of patients with lesions that intact frontal cortex, in particular the medial frontal region, is necessary for good performance of mentalising tasks. Rowe et al. (2001) found that patients with frontal lesions had difficulties with mentalising tasks and these difficulties were independent of problems they had with traditional executive tasks. Stuss et al. (2001) also found that patients with frontal lesions had difficulties with mentalising tasks. In particular patients with medial frontal lesions (especially on the right) were unable to detect the deception of a protagonist, a classic task requiring mental state attribution. As yet there is no information about the effects of lesions in circumscribed regions of paracingulate cortex.

From imaging studies it is not possible to locate regions of activation in precise Brodmann areas. In terms of Talairach co-ordinates the activations observed in the present and previous studies cluster around the border between area 32 and area 9 in paracingulate cortex. The paracingulate (BA32) is cytoarchitectonically
distinct from the anterior cingulate proper and has been described as a cingulo-frontal transition area (Devinsky et al., 1995). The anterior cingulate is a large region and there have been a number of recent reviews concerned to describe functional subdivision within it (e.g. Picard & Strick, 1996; Bush et al., 2000; Barch et al., 2001). The most posterior part of it (the cingulate motor area or caudal cingulate zone) is part of the motor system and, in the monkey at least, sends projections directly to spinal cord (Picard & Strick, 1996). The region of ACC immediately anterior to this (posterior rostral cingulate zone) tends to be activated in executive tasks like the Stroop (Barch et al, 2001) and is sometimes referred to as a cognitive region. The most inferior regions of anterior cingulate (BA 24, 25 and 33) seem to have a role in the processing of emotion (Bush et al., 2000). The region activated in the present study and in previous studies of mentalising lies just anterior to the most anterior part of anterior cingulate proper between the putative cognitive and emotional divisions (see Figure 4-2).

Previous neuroimaging studies have attributed activation in regions of anterior cingulate cortex to autonomic arousal, in particularly cognitive uncertainty and anticipatory arousal (Critchley et al. 2000; Critchley et al., 2001). It could be argued that in the present study the critical activation was due to increased anxiety experienced during the mentalising condition. However, the region activated in those studies was significantly more posterior (Talairach coordinates, R= 8,28,24 and L= -6,28,20) to the anterior paracingulate cortex seen here (R= 8,54,12 and
Furthermore the current study was controlled so that during the scanning window the only actual difference between the two main conditions was in what the subjects believed they were doing. Thus, performance was at chance during both conditions and any anxiety generated from the resulting feedback should be the same.

It might be argued that the paracingulate activity seen here results from an emotional state evoked by the volunteer interacting with another person. In previous imaging studies of mentalising, comparison conditions acting as non-mentalising controls have included interactions between people (Fletcher et al., 1995a; Chapter 3; Brunet et al., 2000; Castelli et al., 2000) which have failed to activate the paracingulate cortex and thus in these instances social emotion has been ruled out. Furthermore, the area activated in this and previous studies of mentalising ability is anterior and superior to the region that Bush et al. (2000) consider to be involved in emotion.

One final argument might be that the activity observed in paracingulate cortex was a result of differences in task difficulty. Mentalising tasks may be intrinsically more difficult and involve components typical of executive tasks. This also seems unlikely. In a meta-analysis of functional imaging studies in which cognitive demand was manipulated, Duncan and Owen (2000) found a
dorsal part of the anterior cingulate cortex to be associated with increasing task difficulty. However the coordinates of the anterior paracingulate activations seen in this and other imaging studies of mentalising are more anterior to this region. Like-wise in the meta-analysis of ACC activations associated with Stroop-like tasks reported by Barch et al. (2001) the focus of activity was posterior (Talairach co-ordinates 3, 19, 35) to that observed in studies of mentalising. Finally, in the current study anterior paracingulate activity was seen when mentalising was compared with both the rule solving and baseline (random generation) conditions. Thus, if this activity were associated with performance anxiety, emotional reward or task difficulty one would expect to see at least intermediate activity in this region in the rule solving v baseline (random selection) comparison. This was not the case.

There are, however, a number of studies that did not require volunteers to make attributions about the mental states of others that nevertheless did activate a similar region of paracingulate cortex (Frith, 2001). These studies have involved some form of self-monitoring; visual self-recognition (Kircher et al., 2000), memory for autobiographical events (Maguire and Mummery, 1999; Maguire et al., 2000), monitoring one’s own speech (McGuire et al., 1996a.), monitoring thoughts (McGuire et al., 1996b.), perception of tickling (Blakemore et al., 1998), perception of pain (Rainville et al., 1997), and reporting emotional responses (Lane et al., 1997). Taken together these results suggest that paracingulate cortex
is active when we make attributions about our own mental states as well as those of other people (see Vogeley et al., 2001 for a discussion of the differences between making attributions about the self and others).

4.5 Conclusions

On the basis of the present and previous studies it may be concluded that paracingulate cortex has a special role in the ability to use information to make attributions about mental states, i.e. adopting an intentional stance. This study has demonstrated that this region is responsible for mentalising in the absence of explicit behavioural information.

Chapters 3 and 4 have used functional imaging to examine directly the neural correlates of mentalising ability in normal adults. The next chapter investigates the use of mentalising skills during a component of “everyday” social behaviour; the recognition of communicative hand gestures. The study was performed on both normal children and children with autism in order to examine both normal and impaired development of this ability.
5.1 Introduction

Chapters 3 and 4 have directly investigated the neural correlates of ‘theory of mind’ ability using fMRI and PET imaging techniques on normal adults. The current chapter uses a different approach to understanding theory of mind, in more “real life” circumstances. It discusses a behavioural study of normally developing and autistic children and their ability to represent two different categories of communicative gestures, instrumental and expressive. Instrumental gestures are defined as gestures that are intended to regulate or change the behaviour of others by communicating commands, e.g. “go away”, “be quiet”, “look over there”. Expressive gestures, on the other hand, are gestures deliberately expressing inner feeling states, or responses to feeling states in others e.g. “I’m embarrassed”, “I’m happy”, “I’m angry” (Barten, 1979). It has been proposed that the comprehension of expressive gestures presupposes a concept of mental states; one’s own as well as those of other people. Instrumental gestures can be understood without this concept as through such gestures one attempts to directly influence another’s behaviour, in the same way as one would manipulate an object in order to obtain a simple physical effect.
As discussed earlier, in chapter 1, autism is a biologically based developmental disorder. Children with this disorder demonstrate impairments in a triad of behavioural abilities. These are socialisation, imagination and communication (Wing and Gould, 1979). It has been proposed that the co-occurrence of these three handicaps can be accounted for in an impairment of “theory of mind” ability or “mind-blindness” (Happé and Frith, 1996). This has been demonstrated with a number of specially developed tests that compare two behaviours differing only in the demands they make upon the child’s ability to mentalise. Examples are: the sabotage/deception task (Sodian and Frith, 1992), the false photograph/false belief test (Leekham and Perner 1991, Leslie and Thaiss 1992), understanding see/understanding know (Perner et al. 1989) and understanding literal/metaphorical expression (Happé, 1993). All these tests make fine cut distinctions between intact and impaired abilities by using cognitively similar tasks. Autistic individuals are impaired on each of the tasks involving an understanding of mental states but not it’s cognitive equivalent. Thus it is argued that the current test is another “fine cuts” comparison in which the ability to recognise expressive gestures differs from the ability to recognise instrumental gestures only by the requirement to understand the minds of others.

Earlier studies by Hobson (1986a; 1986b) examined how well autistic children can comprehend feeling states when they are conveyed through facial expressions, bodily gestures, or voice. The children were shown schematic drawings of faces or gestures expressing pleasure, sadness, fear or anger and
asked to point to the face or gesture most appropriate to a vocalisation, action, or situational context presented on audio tape or video. The autistic children were strikingly impaired on these tasks, relative to the retarded non-autistic controls, suggesting a failure in appreciation of expressive gestures in autistic children. Attwood et al. (1988), by observing children in their school environment, demonstrated that autistic children do not spontaneously produce and respond to expressive gestures but show no such impairment with instrumental gestures.

The current study was divided into two parts. Part 1 (5.2) describes an initial study conducted to establish the developmental age at which normal children are able to represent gestures and also expose a potential difference between the two categories of gesture; instrumental and expressive. Part 2 (5.3) describes a similar study but with children diagnosed with autism. The aim of this study was to establish that the cause of the autistic child’s failure to produce and respond to expressive gestures is a failure to recognise these gestures, consistent with the notion of impaired theory of mind ability. It was hypothesised that divergent or dissociable neural networks exist for representing these two categories of gesture, parts of which may be selectively damaged, resulting in the deficits manifest in autistic individuals. This cannot be directly tested in this study, but may or may not be reflected in differences in the developmental age at which normal children acquire the ability to understand these two types of gesture. Also, based on the findings of Attwood et al. (1988) it was predicted that the autistic children would be unable to recognise the expressive gestures.
5.2 Part 1

5.2.1 Methods

5.2.1.1 Subjects

60 normal children (29 male and 31 female) with no known history of neurological or psychiatric disorders or developmental delay were tested on their ability to represent gestures. The children, all of whom attended either primary school or nursery, were aged between 4 years and 7 years. They were divided into three groups of twenty: 4-5 years (mean = 4.6; SD = 0.17); 5-6 years (mean = 5.5; SD = X) and 6-7 years (mean = 6.3; SD = X). Informed parental consent was obtained for all the participating children prior to testing. This study was approved by the ethics committee of University College London (UCL).

5.2.1.2 Task

The children were presented a total of 46 silent video clips of duration 2-3s. Each video showed the experimenter miming either an expressive gesture or an instrumental gesture (see table 5-1 for full list). The child was asked after each gesture for the meaning. The child’s answer was reported on a score sheet as correct or incorrect. Meanings or explanations that did not match the score sheet but were deemed appropriate were scored as correct (e.g. a shrug meaning “I don’t know” could also be interpreted as meaning, “I don’t care”). Expressive gestures to which instrumental meanings were attributed, or vice versa, were
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

scored as incorrect. The video was repeated more than once if the child’s attention strayed during the first showing, or if the child requested it be shown again.

<table>
<thead>
<tr>
<th>Table 5-1. List of the 23 expressive and 23 instrumental gestures.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expressive</strong></td>
</tr>
<tr>
<td>Wave a greeting</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Well done</td>
</tr>
<tr>
<td>Impatience</td>
</tr>
<tr>
<td>I don’t care</td>
</tr>
<tr>
<td>Fear</td>
</tr>
<tr>
<td>Boredom</td>
</tr>
<tr>
<td>I feel cold</td>
</tr>
<tr>
<td>I feel hot</td>
</tr>
<tr>
<td>That’s naughty</td>
</tr>
<tr>
<td>Excitement</td>
</tr>
<tr>
<td>I don’t know</td>
</tr>
</tbody>
</table>

The gestures were filmed using a digital camera (Sony, DCR-PC1E) and edited using DVgate Motion Version 1.4.00 (Sony) and Adobe Premier (version 6) software then converted into avi (audio video interleaved) format. Gestures were presented in a pseudorandomised order on a laptop PC (Sony Vaio) in Microsoft® PowerPoint® 97 SR-1.

5.2.1.3 *Data analysis*

Data were analysed using SPSS (version 9). Differences between the two categories of gesture within each age group were determined with matched sample two-tailed *t* tests. Differences in each category between age groups were
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

determined with both independent sample two-tailed \( t \) tests and Kruskal-Wallis 1-way ANOVA. The latter analyses were performed so that a consistent comparison could be made with the results of Attwood et al. (1988) in which the sample size varied across the groups and the number of levels was much smaller, thus an independent sample \( t \) test would have been inappropriate.

5.2.2 Results

Table 5-2 shows the mean scores of each age group of children on instrumental and expressive gesture recognition. The children significantly improved their performance on both types of gesture with increasing age (see figure 5-1) as indicated by both independent sample two-tailed \( t \) tests and Kruskal Wallis 1-way ANOVA. This improvement was most significant between the ages 4 and 5 (\( t (38)=5.5, p<0.001 \) instrumental; \( t (38)=5.8, p<0.001 \) expressive) \( (\chi^2(1)=17.5, p<0.001 \) instrumental; \( \chi^2(1)=9.8, p=0.002 \) expressive) (comparison of 5 and 6 year olds – \( t (38)=4.1, p=0.001 \) instrumental; \( t (38)=3.3, p=0.002 \) expressive) \( (\chi^2(1)=19.6, p<0.001 \) instrumental; \( \chi^2(1)=11.0, p=0.001 \) expressive)). There was no significant difference, within any of the age groups, in the children’s’ ability to recognise instrumental and expressive gestures (figure 5-1) as indicated by matched sample two-tailed \( t \) tests. Further independent sample two-tailed \( t \) tests showed that there were no differences in the differences between the two categories, for any of the age groups.
Table 5.2. Representation of instrumental and expressive gestures in normal children.

<table>
<thead>
<tr>
<th>Group</th>
<th>Age range</th>
<th>Instrumental (max = 23)</th>
<th>Expressive (max = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>6-year olds</td>
<td>6.1-6.5</td>
<td>6.27</td>
<td>0.8</td>
</tr>
<tr>
<td>5-year olds</td>
<td>5.1-5.9</td>
<td>5.55</td>
<td>2.24</td>
</tr>
<tr>
<td>4-year olds</td>
<td>4.3-4.9</td>
<td>4.6</td>
<td>2.67</td>
</tr>
</tbody>
</table>

Figure 5-1. Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of the 4-year-old, 5-year-old and 6-year-old groups of normal children.

5.2.3 Discussion

The aims of this experiment were to: 1) establish the developmental age at which normal children are able to represent gestures and 2) expose a potential difference between the two categories of gesture, instrumental and expressive. It has been
argued that the recognition of expressive gestures (gestures deliberately expressing inner feeling states or responses to feeling states in others) requires a concept of the mental states of others. This is not necessary for the recognition of instrumental gestures (gestures intended to regulate or change the immediate behaviour of others by communicating commands) (Barten, 1979). This suggests divergent neural pathways exist for the representation of instrumental and expressive gestures.

With regard to the first aim, the results show that the children clearly improved their performance with increasing age. This difference was significant across all groups but was most significant between the ages of 4 and 5 (figure 5-1). This is consistent with the findings of Attwood et al. (1988) that reported significant differences between the four and five year old groups only, on their performance of recognising instrumental gestures. In the Attwood study (Attwood et al., 1988) only eight very common gestures were considered, compared with 23 (of each category) in this study. In the current study it was often the same gestures that the five and six-year-old group failed to recognise. These were: “Nervous”, “Impatience”, “Pleading”, “Thoughtful”, “Embarrassment” (expressive) “Think”, “I surrender”, “Don’t be nosey”, “Hurry up”, “Kiss me on my cheek” (instrumental). This suggests that experience plays an important role in the recognition of the more sophisticated gestures as older children have had longer to learn these meanings as a result of their increased exposure to these gestures.
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

There was one other notable difference between the responses of the 4-year-old group and the five- and six-year-old groups. When the older groups did not know the meaning of a gesture they were more likely to respond “I don’t know”. The younger children by contrast, were more inclined to attribute a (mistaken) meaning. They would respond, for example, “she’s pointing to her cheek” “she’s playing” or simply “hand”, “cheek” or “hair”. These results indicate that while normal children have developed the ability to recognise several commonly used gestures at the age of 4 years, they become significantly more proficient at this by 5 years.

In relation to the second aim, the results demonstrated no significant differences in normal children’s ability to recognise expressive compared with instrumental gestures. The children tested in the study were aged four or over. Although no measure of theory of mind ability was acquired, it is reasonable to assume that these children had an adequate concept of mind to pass first-order false belief tasks (Wimmer and Perner, 1983). Therefore it is possible that a younger group of children may have shown a difference in the two categories. However, the current paradigm is inappropriate for children with verbal ability below that of a four-year-old. The previous study of gesture comprehension in normal 3-4 year olds, involved testing the child’s ability to respond appropriately to a limited series of commonly used gestures (Attwood et al., 1988). This was not suitable for the current study as a larger selection of gestures was examined; many of which do not necessitate a response. This is particularly pertinent to the expressive gestures.
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

While a result of a significant difference would have been evidence to support the idea of dissociable or divergent neural pathways for the representation of these two categories of gesture, this negative result does not rule out such a notion. Instead it indicates that between the ages of 4 and 6, the abilities to interpret both instrumental and expressive gestures develop in parallel in normal children.

5.3 Part 2

5.3.1 Methods

5.3.1.1 Subjects

10 children/adolescents, all male, with the diagnosis of autism were also tested on their ability to represent gestures. The children, all of whom attended specialist schools, were aged between 8.3 years and 16.3 years. Only children with at least some spontaneous speech were included in the study. Eight schools were approached to participate. The sample was taken from four of these. Out of the three schools remaining, two had no pupils with adequate verbal mental age (VMA) and one school asked not to be included because of other commitments. Informed parental consent was obtained for all the participating children prior to testing. This study was approved by the ethics committee of University College London (UCL).
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

5.3.1.2 Tasks

The children were presented with the same task as in part 1 (5.2.1.2). In addition the autistic individuals were tested for verbal mental age with the British Picture Vocabulary Scale (BPVS II) (Dunn et al. 1997) test.

5.3.1.3 Data analysis

Data were analysed using SPSS (version 9). Differences between the two categories of gesture within the group were determined with matched sample two-tailed t tests. For further comparison and to eliminate verbal mental age as a confound, the same tests were performed with the group divided into those with a verbal mental age > 6 and those with a verbal mental age < 6. The results of the autistic children were not directly compared with age-matched or verbal I.Q matched normal controls from part 1 (5.2). However, differences in the difference scores of the two categories of gestures were examined between all groups with independent sample two-tailed t tests.

5.3.2 Results

Table 5-3 and Figure 5-2 show the mean scores of the autistic group of children on instrumental and expressive gesture recognition. A matched sample two-tailed t test indicates a significant difference in the ability to recognise instrumental and expressive gestures with the group being impaired on the latter category (t (9)=13.1, p< 0.001). This effect was also evident when the autistic group were
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

divided into those with a verbal mental age greater than 6 (t (3)= 5.5, p=0.012) and those with a verbal mental age less than 6 (t (5)= 20, p< 0.001) (see figure 5-3).

Independent sample two-tailed t tests demonstrated significant differences in the difference scores of the two categories of gesture when the autistic group were compared to the normal children (Comparison with: 4-year-olds, t (28)=5.2, p < 0.001; 5-year-olds, t (28)=9.6, p < 0.001; 6-year-olds, t (28)=-9.6, p < 0.001).

Table 5-3. Representation of instrumental and expressive gestures in children with autism.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Verbal mental age</th>
<th>Age range</th>
<th>Instrumental (max=23)</th>
<th>Expressive (max = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4.07-5.01</td>
<td>8.3-13.4</td>
<td>9.5</td>
<td>2.87</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6.04-12.1</td>
<td>11.5-16.3</td>
<td>16.15</td>
<td>1.95</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>4.07-12.1</td>
<td>8.3-16.3</td>
<td>13.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

Figure 5-2. Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of each group.
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

Figure 5-3. Graph showing the mean performance scores and standard deviations, for instrumental and expressive gestures, of the two autistic groups (verbal mental age < 6 and verbal mental age > 6) and the 4, 5 and 6-year-old groups.

Discussion

The results of this study show that the autistic children are significantly impaired on their recognition of expressive gestures compared to their ability to recognise instrumental gestures. This is in contrast to the normal children who showed that from the age of 4 years old these abilities develop in parallel. The results show a significant difference in the difference of the autistic group’s performance on expressive and instrumental gestures compared with that of the three groups of normal children. This is regardless of verbal mental age as, in contrast to the controls, all the autistic volunteers show this discrepancy in their performance.
Earlier work by Attwood et al. (1988) described a series of studies aimed at examining autistic children's understanding and use of interpersonal gestures. They showed that autistic children are able to correctly respond to instrumental gestures and initiate them on verbal request. They also observed that autistic children will spontaneously produce instrumental gestures with the same frequency as Down's syndrome adolescents but they never produce expressive gestures. They argued that this failure to produce and respond to expressive gestures is a failure to recognise these gestures, consistent with the notion of impaired theory of mind ability.

It is clear, from these results that the autistic children have more difficulty in recognising gestures which influence mental states than gestures which manipulate behaviour. Similarly, Baron-Cohen (1989) found that autistic children were impaired in their use and understanding of pointing for the sake of sharing attention ("protodeclaritive pointing") but not for pointing in order to obtain a desired object ("protoimperative pointing"). This study provides further evidence to support Happé and Frith's psychological theory (Happé and Frith, 1996) aimed at explaining the triad of impairments in autism in terms of a single underlying cognitive deficit – a failure or delay in mentalising.

The autistic children's ability to recognise instrumental gestures was shown to be only slightly worse than would be expected from verbal mental age, although this was not tested statistically here. The autistic group, which was much smaller than
the groups of normal children, consisted of children with a range of chronological and verbal ages and performance within the group varied considerably. Therefore it was deemed inappropriate to divide them into comparative groups for statistical analysis. However, figure 5-2 demonstrates that on average the autistic group’s performance on the instrumental gestures was comparable with the normal 4-year-old group. The more verbal and generally older autistic children showed a similar level of performance to the normal 5-year-old group (see figure 5-3). The autistic children with a verbal mental age of less than 6 performed slightly worse than the normal 4-year-old group. The autistic children’s ability to recognise expressive gestures was shown to be much worse than would be expected from verbal mental age. Figure 5-3 shows that both groups of autistic children were worse than the normal 4-year-old group. This effect was much more extreme in the children with a verbal mental age less than 6.

However, the autistic group showed that their recognition of expressive gestures (as with instrumental gestures) improves with increasing verbal mental age indicating that this ability can be acquired. Whether this is coincident with an improvement in theory of mind ability or indicative of the employment of another strategy/mechanism remains unclear. A relationship between theory of mind ability and language acquisition has been examined in previous studies of autistic children. For instance, it has been proposed that a theory of mind deficit may hamper word learning in the developing child, by interrupting processes of joint attention, reference and ostention (Frith and Happé, 1994). Thus, the BPVS
Chapter 5 Representation of expressive and instrumental gestures in normal & autistic children.

(Dunn et al., 1997) used in this study, which is a measure of the number of words known, may inadvertently be a measure of theory of mind ability in the children with autism. However, Happé (1995) has shown that children with autism require a far higher verbal mental age than normal children to pass standard theory of mind tests. While normally developing children had a 50% probability of passing false belief tasks at the verbal mental age of 4 years, autistic children took more than twice as long to reach this probability of success, at the advanced VMA of 9.16. Happé (1995) showed that some minimal mental age level is necessary for theory of mind success (VMA = 5.5) and that there is a higher verbal mental age (VMA = 11.58) above which all volunteers pass false belief tasks.

One interesting interpretation of the high VMA found in autistic who pass theory of mind tasks is to suggest that they are solving the tasks in a verbally mediated fashion (Happé, 1995). Based on the conclusions drawn from the Attwood study (Attwood et al., 1988) it was hypothesised that, in the current study, autistic children would be unable to recognise any of the expressive gestures. This was clearly not the case. Even some of the children with lower verbal mental ages were able to recognise a few. Many of the same expressive gestures were recognised by each child. These were: “I’m cold”, “I’m “hot”, “I’m full”, “I like it” (thumbs up sign) “I don’t like it” (thumbs down sign) “well done” (clapping hands) and “wave a greeting”. It is possible that these gestures are commonly used in the school environment and verbally reinforced. In this circumstance it is feasible that these gestures are learned and processed in a similar manner to the
instrumental gestures. All of the children in this study attended one of three specialist schools. Two of these schools claimed to spend some time training the children on social skills, including recognising facial expressions such as happy, sad and angry. Indeed one of the more verbal children responded to the happy gesture with the words "she is smiling so she must be happy". Thus it is feasible that, for the autistic children the route to recognising some of these gestures is strategic and explicit i.e. through a process of deduction, rather than automatic and implicit which is thought to be the case with normal children and adults. Indeed it is believed that the small minority of autistic individuals who are able to pass standard false belief tasks do so with a learned and general reasoning process rather than the automatic and innate theory of mind mechanism.

5.4 General discussion

The results from these two experiments show that normal children develop the ability to recognise both expressive and instrumental gestures by the age of 5. Also, from 4 years onwards at least, this development observes a similar time course. However, the results from part 2 show that the autistic children demonstrate an impairment in their ability to recognise expressive gestures. This is consistent with earlier studies which examine "fine cuts" between similar cognitive behaviours, in which children with autism show distinct impairments on those behaviours which require a concept of mental states (Attwood et al., 1988; Baron-Cohen 1989; Perner et al., 1989; Baron-Cohen, 1992). These results
suggest that divergent or dissociable neural pathways exist for the perception and recognition of these two categories of gesture, which may be selectively damaged, resulting in the deficit manifest in autistic individuals. Furthermore, if this failure by the autistic children to “appreciate” expressive gestures is borne out of an inability to represent mental states then it is reasonable to expect that the neural network for representing such gestures will involve regions of the brain associated with such an ability.

Chapter 7 describes a functional neuroimaging study examining the neural correlates of the perception and recognition of expressive and instrumental gestures which, based on the results of the current study, are divergent. It is hypothesised that the recognition of expressive gestures will involve areas of the brain associated with theory of mind processing. Chapters 3 and 4 demonstrated that theory of mind ability is mediated by a highly circumscribed region of the brain – the anterior paracingulate cortex (BA 32/8/9). This is consistent with the results of previous imaging studies (Fletcher et al., 1995; Goel et al., 1995; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001, McCabe et al., 2001). In the majority of these studies (Fletcher et al., 1995; Goel et al., 1995; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001) two other regions have been consistently activated. These are the posterior part of the right superior temporal sulcus (STS) at the temporo-parietal junction and the temporal poles bilaterally. However, these regions are not necessarily associated with mentalising alone. A previous functional imaging study of
biological motion (Bonda et al., 1996) reported activity in all three of these regions when they compared expressive body movements with goal-directed hand actions. This suggests that such a network exists which includes the now well established theory of mind region – the anterior paracingulate cortex. A major aim of chapter 7 is to explore the hypothesis that the recognition of expressive, but not instrumental, gestures engages a network of brain regions involved in theory of mind. This would account for the impaired performance of this task in autistic children and improve our understanding of the biological basis of this disorder.
6 Chapter 6: The Neural Correlates of Gesture Representation: Expressive and Instrumental.

6.1 Introduction

Chapter 5 discussed the contribution of theory of mind to our everyday ability to recognise communicative gestures. Two types of gesture were examined, expressive and instrumental. Expressive gestures are gestures that express feeling states. It is believed that the recognition of such gestures requires a theory of mind. In contrast instrumental gestures are gestures intended to regulate or change the behaviour of others by communicating commands. This study demonstrated that individuals with autism, who show a deficit in theory of mind abilities, are able to recognise instrumental gestures but are impaired on their recognition of expressive gestures. This result indicates that distinct neural pathways exist for the processing of these two categories of gesture. The current chapter uses fMRI to examine the neural correlates involved in the representation of expressive and instrumental gestures. It was hypothesised that neural regions associated with mentalising ability would be activated during the recognition of expressive gestures.

Chapters 3 and 4 of this thesis have already described two quite different studies aimed at characterising the brain regions involved in mentalising ability. Chapter 4 demonstrated that in the absence of explicit behavioural cues the anterior
paracingulate cortex alone mediates this ability. However, in addition to anterior paracingulate cortex, Chapter 3 activated a number of other brain regions in association with theory of mind tasks. Two of these regions, the superior temporal sulcus (STS) (particularly on the right) and the temporal poles bilaterally, are consistently activated in previous neuroimaging studies of theory of mind ability (Fletcher et al., 1995; Goel et al., 1995; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001). In most of these studies (Fletcher et al., 1995; Goel et al., 1995; Chapter 3; Brunet et al., 2000; Castelli et al., 2000) the activation in the STS has been attributed to the involvement of people or characters rather than mentalising *per se*. This is because the same region has activated in both mentalising and equivalent non-mentalising conditions (Brunet et al., 2000; Castelli et al., 2000). Brunet et al., 2000 found right superior temporal sulcus (STS) to be associated with the involvement of people compared to physical objects alone and Castelli et al. (2000) also found right STS associated with attributing intentions to, but not necessarily mental states, to animated geometric shapes.

The superior temporal sulci have activated in previous neuroimaging studies which have sought to characterise the neural correlates of action observation (Decety et al., 1997; Grezes et al., 1998; Grezes et al., 1999) and the perception of biological motion (Bonda et al., 1996; Calvert et al., 1997; Puce et al., 1998; Grossman et al., 2000). Similarly studies of single-cell recordings in monkeys (Oram and Perrett, 1994) have found cells in the STS which respond to moving
hands and faces, but not movement of inanimate objects. The experiment by Bonda et al. (1996) is of particular interest to the current study and forms the basis of the apriori hypothesis, as they found activity in a number of other brain regions which have been associated with mentalising in previous neuroimaging studies (Fletcher et al., 1995; Goel et al., 1995; Chapter 3; Brunet et al., 2000; Castelli et al., 2000) including the medial prefrontal cortex (BA 9), the right superior temporal sulcus and the temporal poles bilaterally. They used PET to study the perception of biological motion with point light displays by comparing goal directed hand actions and expressive whole body motion with object motion and random motion. They reported activations in the rostro caudal part of the right superior temporal sulcus and adjacent temporal cortex, the amygdala bilaterally and the left medial prefrontal cortex (BA 9) to be associated with the perception of signs conveyed by expressive body movements. While goal directed hand actions (i.e. imitating the act of reaching toward a glass, picking it up and bringing it to the mouth) activated regions of the left hemisphere in the intraparietal sulcus and caudal superior temporal sulcus. These results indicate that while the STS is involved in the perception of biological motion in general, the right STS is particularly sensitive to expressive actions involving mentalising processes.

As mentalising is a capacity confined to humans and possibly some chimpanzees (Heyes, 1998) it has been suggested that this ability has evolved from relevant preexisting abilities. These include; the ability to distinguish between animate and
inanimate entities, the ability to share attention by following the gaze of another agent, the ability to represent goal directed actions and the ability to distinguish between actions of the self and of others. It has been hypothesised that the right STS is involved in the “detection of the behaviour of agents and analysis of the goals and outcomes of this behaviour” (Frith and Frith, 1999).

The main aim of the current chapter is to test the hypothesis that the right superior temporal sulcus is involved in a network of social perception that contributes to, but is not essential for, mentalising ability. This region may be a necessary prerequisite for the development of this ability in children (Frith and Frith, 1999). This study sought to determine the neural networks involved in the representation of instrumental and expressive gestures and reveal a correlation between the anterior paracingulate cortex (BA 9/32) which is thought to mediate mentalising ability and the right superior temporal sulcus at the temporo-parietal junction. It was hypothesised that the representation of expressive gestures would engage a network of regions relevant to mentalising, including the right temporo-parietal cortex, the medial prefrontal cortex (BA 9/32) and the temporal poles, while representation of instrumental gestures would engage different brain regions predominantly in the left hemisphere reflecting language processing.
6.2 Methods

6.2.1 Subjects
Twelve right-handed volunteers with no neurological or psychiatric history participated in this study. Of these, eight were male and five were female, with a mean age of 35.98 (28.4 - 59.5 years). This study was approved by the ethics Committee of the Institute of Neurology, UCL. Informed written consent was obtained from all volunteers prior to scanning.

6.2.2 Tasks
All stimuli were displayed on a monitor and presented to the subject via a 45° angled mirror positioned above the head coil; this mirror was adjusted to be within the subjects field of vision without having to tilt his/her head. A test image was presented on the screen prior to scanning to ensure that the image was in focus and the subject could comfortably read the text.

Stimuli consisted of 44 silent video clips of 3-second duration (See table 6-1 for full list of stimuli). Each video depicted one of two people miming a gesture. 22 depicted expressive gestures. These are gestures deliberately expressing inner feeling states, or responses to feeling states in others e.g. embarrassment, excitement or anger. 22 depicted instrumental gestures. These are gestures that are intended to regulate or change the behaviour of others by communicating commands, e.g. go away, be quiet, look over there. A null video was constructed
from the first frame of one of the gestures, showing the experimenter in a neutral position. This frame was repeated continuously for three seconds, thus there was no movement. Faces were blurred on all stimuli to rule out facial expression as a potential confound.

The gestures were filmed using a digital camera (Sony, DCR-PC1E) and edited using DVgate Motion Version 1.4.00 (Sony) and Adobe Premier (version 6.0) software. New videos were produced for this study as those presented in the earlier behavioural study (Chapter 5) were poorly controlled for visual content. For instance, the distance between the camera and the actor and therefore the magnification varied, and the field of view in relation to the actor's body length differed depending on the gesture. In the current study the camera to actor distance and lighting levels were kept constant and the same portion of the actor was viewed in each movie, this was an area between the upper thigh and approximately 1ft above the head of the actor. Faces were blurred using Adobe After Effects (version 4.0) then each clip was saved as a QuickTime movie. The videos were presented to the volunteers via an Apple Macintosh computer.

6.2.3 Data Acquisition

A Siemens VISION MRI system operating at 2 Tesla was used to acquire both T1 weighted anatomical and echo-planar T2* weighted image volumes with blood oxygenation level-dependent (BOLD) contrast. Functional images were acquired in one continuous run; the order of the conditions was counterbalanced across
Each image volume constituted 36 3mm axial slices with in-plane resolution of 3mm by 3mm positioned to cover the whole brain. Volumes were acquired continuously every 2736 ms while subjects performed five experimental conditions: 1) observing expressive gestures, 2) observing instrumental gestures, 3) observing and recognising expressive gestures, 4) observing and recognising instrumental gestures 5) null condition. Each task epoch comprised 1.55 image volumes and was 4 seconds duration; a 3-second video plus a 1-second fixation.

The gestures were blocked into groups of 6 epochs. These blocks defined the experimental condition; either “observe only” or “observe and recognise”. Each block consisted of a selection of expressive or instrumental gestures and null events, in a pseudorandomised order. Each gesture video was presented twice. The repeated video was always presented under the same experimental condition; either “observe only” or “observe and recognise”. Thus each study consisted of a total of 264 epochs, 176 from each of the four main experimental conditions and an additional 88 null epochs. 44 repeats of the null video were presented in “observe only” blocks and 44 in “observe and recognise” blocks.

Prior to each block “Get Ready” was displayed on the screen for 1 second followed by a question which defined the experimental condition for that block for a duration of 3 seconds. The question was either “Which hand is higher?” denoting an “observe only” block or “Is the gesture positive?” denoting an “observe and recognise” block. During an “observe only” block, responses were
either “right”, “left” or “both”. Volunteers were instructed to decide which hand had reached the highest point at the end of the video. During an “observe and recognise” block, responses were either “yes”, “no” or “neutral”. In this instance they had to decide if the gesture was welcoming, unwelcoming or neutral.

Following each video and null video a fixation cross was displayed for a second. The volunteers were asked to make their response after each video when the cross was displayed. Responses were made on an MRI compatible keypad with three keys. Volunteers were instructed to make no response to the null events.

Each functional imaging run began with five volumes that were discarded prior to analysis to allow for T1 saturation effects. A total of 485 image volumes were acquired of which 480 were analysed. The duration of the scanning was approximately 45 minutes.

<table>
<thead>
<tr>
<th>Table 6-1. List of the 22 expressive and 22 instrumental gestures.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expressive</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Wave a greeting</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Well done</td>
</tr>
<tr>
<td>Impatience</td>
</tr>
<tr>
<td>I don’t care</td>
</tr>
<tr>
<td>Fear</td>
</tr>
<tr>
<td>Boredom</td>
</tr>
<tr>
<td>I feel cold</td>
</tr>
<tr>
<td>I feel hot</td>
</tr>
<tr>
<td>That’s naughty</td>
</tr>
<tr>
<td>Excitement</td>
</tr>
<tr>
<td>Look down</td>
</tr>
<tr>
<td>Stand up</td>
</tr>
<tr>
<td>Sit down</td>
</tr>
<tr>
<td>Come here</td>
</tr>
<tr>
<td>Go away</td>
</tr>
<tr>
<td>Stop</td>
</tr>
<tr>
<td>Back off</td>
</tr>
<tr>
<td>Be quiet</td>
</tr>
<tr>
<td>Give it to me</td>
</tr>
<tr>
<td>Turn around</td>
</tr>
<tr>
<td>No thank you</td>
</tr>
</tbody>
</table>

136
6.2.4 Data Analysis

Data were analysed using Statistical Parametric Mapping (SPM97, Wellcome Dept. of Cognitive Neurology, London, UK) implemented in MATLAB (Mathworks Inc., Sherborn, MA, USA) and run on a SPARC workstation (Sun Microsystems Inc., Surrey, UK). The imaging time series was realigned using the first image and spatially normalised to the stereotactic space of Talairach and Tournoux (1988) using MNI templates (Montreal Neurological Institute). These data were subsequently smoothed with an isotropic Gaussian kernel of 8mm at full width half-maximum.

Analysis was carried out using the general linear model and a delayed boxcar waveform. Subject-specific low-frequency drift in signal was removed by a high pass filter and global signal changes were removed by including a global covariate (Holmes et al., 1997). A random-effects analysis, which takes account of within and between subject variance, was performed. Effects at each voxel were estimated and regionally specific effects were compared using linear contrasts. The resulting set of voxel values for each contrast constituted a statistical parametric map of the t statistic (SPM\{t\}) which was subsequently transformed to the unit normal distribution, SPM\{Z\}. The statistical threshold used to report activations was set at p<0.001 uncorrected for regions of interest that were specified apriori. The regions in which responses to expressive gestures were specifically predicted apriori are the anterior paracingulate cortex (BA 32/9) and the right superior temporal sulcus (BA 21/37) (Bonda et al., 1996; Chapter
3). This hypothesis is derived from the study by Bonda et al. (1996) who examined a similar comparison of expressive body movements with goal directed hand actions and activated these two regions. Also, from the results of chapter 3, which found activity in the anterior paracingulate cortex and the right and left superior temporal cortex (but more significantly on the right) in association with mentalising tasks. A correction for multiple comparisons was performed within each region of interest using the theory of Gaussian Random fields (Friston et al., 1994), and those voxels which survived correction at p< 0.05 are indicated. Voxels within a region of interest were selected using a binary mask drawn over the anatomically defined boundaries of the region as delineated by the atlas of Duvernoy (1999). A number of other brain regions were also predicted apriori for both instrumental and expressive gestures, based on the findings of Bonda et al. (1996), these are reported as significant at p<0.001 uncorrected. For all other regions, activations are reported as significant at p<0.05 corrected for multiple comparisons.

The stereotactic coordinates of Talairach and Tournoux, (1988) are used to report the observed activation foci. However, descriptions of the anatomical localisation of the foci were determined using averaged structural MRIs of the group and the atlas of Duvernoy (1999).
6.3 Results

6.3.1 Expressive vs. instrumental (conjunction of observe and recognise contrasts)

A significant activation at $p<0.05$ small volume corrected was seen in the right STS (BA 21/37) (Table 6-2) (Figure 6-1). Also significant activations ($p<0.001$, uncorrected) were seen in the right and left amygdala. There was an apriori hypothesis for all of these regions based on the findings of Bonda et al. (1996). Other regions that activated but did not reach a corrected level of significance and for which there was no apriori hypothesis were the right and left dorsal prestriate (BA 17) cortices and the right medial frontal cortex (orbito-frontal cortex) (BA 11).

Figure 6-1. A statistical parametric map (SPM $Z_j$) as a maximum intensity projection showing the areas where there was greater activation to expressive gestures compared to instrumental gestures (conjunction of observe and recognise contrasts) in the group. Views are from the right, top and behind.

6.3.2 Instrumental vs. expressive (conjunction of observe and recognise)

A significant activation ($p<0.05$ corrected) was seen in the left inferior frontal cortex (Table 6-2) (Figure 6-2). The left middle frontal cortex also activated but did not reach a corrected level of significance.
Figure 6-2. A statistical parametric map (SPM[Z]) as a maximum intensity projection showing the areas where there was greater activation to instrumental gestures compared to expressive gestures (conjunction of observe and recognise contrasts) in the group. Views are from the right, top and behind.

<table>
<thead>
<tr>
<th>Region</th>
<th>Putative Brodmann area</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expressive vs. instrumental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R orbito-frontal cortex</td>
<td>11</td>
<td>2</td>
<td>48</td>
<td>-16</td>
<td>3.37</td>
</tr>
<tr>
<td>R amygdala</td>
<td>22</td>
<td>0</td>
<td>-20</td>
<td></td>
<td>3.24</td>
</tr>
<tr>
<td>R amygdala</td>
<td>16</td>
<td>-4</td>
<td>-28</td>
<td></td>
<td>3.75</td>
</tr>
<tr>
<td>L amygdala</td>
<td>-14</td>
<td>-6</td>
<td>-22</td>
<td></td>
<td>3.69</td>
</tr>
<tr>
<td>R superior temporal sulcus</td>
<td>21/37</td>
<td>48</td>
<td>-44</td>
<td>-4</td>
<td>3.77</td>
</tr>
<tr>
<td>R Dorsal prestriate</td>
<td>17</td>
<td>24</td>
<td>-98</td>
<td>14</td>
<td>3.83</td>
</tr>
<tr>
<td>L Dorsal prestriate</td>
<td>17</td>
<td>-24</td>
<td>-100</td>
<td>6</td>
<td>3.59</td>
</tr>
<tr>
<td><strong>Instrumental vs. expressive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Inferior frontal cortex</td>
<td>44</td>
<td>-42</td>
<td>10</td>
<td>24</td>
<td>4.65</td>
</tr>
<tr>
<td>L Middle frontal cortex</td>
<td>6/8</td>
<td>-42</td>
<td>14</td>
<td>46</td>
<td>4.37</td>
</tr>
</tbody>
</table>

6.3.3 Recognise versus observe (conjunction of instrumental vs. expressive)

A significant activation at \( p < 0.05 \) small volume corrected was seen in the right STS (BA 21/37) (Table 6-3) (Figure 6-3). Other regions that activated but did not reach a corrected level of significance were the superior frontal cortex, the left inferior frontal cortex, the left orbito-frontal cortex, the inferior, middle temporal cortices and poles bilaterally and the cerebellum.
6.3.4 Observe versus recognise (conjunction of instrumental vs. expressive)

Significant activations at p<0.05 small volume corrected, were seen in the right and left medial prefrontal cortices (BA 32/9) (Table 6-3) (Figure 6-4). Other regions that activated but did not reach a corrected level of significance were the right posterior cingulate (and the right angular gyrus (BA 39)).
Table 6-3. Main effects comparison of observe and recognise.

<table>
<thead>
<tr>
<th>Region</th>
<th>Putative Brodmann area</th>
<th>X</th>
<th>Y</th>
<th>z</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conjunctions of expressive and instrumental contrasts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognise vs. observe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L superior frontal cortex</td>
<td>8</td>
<td>-10</td>
<td>26</td>
<td>58</td>
<td>3.63</td>
</tr>
<tr>
<td>L inferior frontal</td>
<td>47</td>
<td>-42</td>
<td>36</td>
<td>-16</td>
<td>3.58</td>
</tr>
<tr>
<td>L inferior frontal</td>
<td>45</td>
<td>-52</td>
<td>26</td>
<td>10</td>
<td>3.40</td>
</tr>
<tr>
<td>L orbito frontal cortex 11/47</td>
<td>11</td>
<td>-26</td>
<td>26</td>
<td>-20</td>
<td>3.62</td>
</tr>
<tr>
<td>R inferior temporal cortex</td>
<td>20</td>
<td>48</td>
<td>20</td>
<td>-26</td>
<td>3.18</td>
</tr>
<tr>
<td>L inferior temporal</td>
<td>20</td>
<td>-46</td>
<td>18</td>
<td>-26</td>
<td>3.37</td>
</tr>
<tr>
<td>R temporal pole</td>
<td>20</td>
<td>48</td>
<td>20</td>
<td>-26</td>
<td>3.18</td>
</tr>
<tr>
<td>L temporal pole</td>
<td>20</td>
<td>-46</td>
<td>18</td>
<td>-26</td>
<td>3.37</td>
</tr>
<tr>
<td>R superior temporal sulcus</td>
<td>21</td>
<td>54</td>
<td>-42</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Cerebellum</td>
<td></td>
<td>34</td>
<td>-80</td>
<td>-38</td>
<td>3.44</td>
</tr>
<tr>
<td>Observe versus recognise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L medial fronto-polar cortex</td>
<td>32/10</td>
<td>-6</td>
<td>48</td>
<td>-4</td>
<td>3.66</td>
</tr>
<tr>
<td>R medial fronto-polar cortex</td>
<td>32/10</td>
<td>16</td>
<td>50</td>
<td>-2</td>
<td>3.48</td>
</tr>
<tr>
<td>R posterior cingulate cortex</td>
<td>31</td>
<td>10</td>
<td>-30</td>
<td>40</td>
<td>3.48</td>
</tr>
<tr>
<td>R angular gyrus</td>
<td>39</td>
<td>44</td>
<td>-68</td>
<td>32</td>
<td>3.67</td>
</tr>
</tbody>
</table>

6.3.5 Interaction: (recognise expressive (RE) – observe expressive (OE)) x (observe instrumental (OI) - recognise instrumental (RI))

Significant activations (p<0.05, small volume corrected) were seen in the left medial prefrontal cortex and the right superior temporal sulci (Table 6-4) (Figure 6-5). A significant activation was also in the left STS (BA 21/37). There was an apriori hypothesis for this region based on the findings of Bonda et al. (1996). Other regions that activated but did not reach a corrected level of significance and for which there was no apriori hypothesis were the right and left medial frontal cortex (BA 8), the right and left temporal poles (BA 21), the right insula and the L precentral gyrus (BA 4).
6.3.6 Interaction: (recognise instrumental (RI) \textendash{} observe instrumental (OI)) \times
(observe expressive (RE) \textendash{} recognise expressive (RE))

No significant activations were seen in this contrast (Table 6-4).

<table>
<thead>
<tr>
<th>Region</th>
<th>Brodmann area</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(OE \textendash{} OE) \times (OI \textendash{} RI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L medial prefrontal cortex</td>
<td>32/9</td>
<td>-12</td>
<td>54</td>
<td>0</td>
<td>3.15</td>
</tr>
<tr>
<td>F medial prefrontal cortex</td>
<td>32/9</td>
<td>14</td>
<td>52</td>
<td>10</td>
<td>2.7</td>
</tr>
<tr>
<td>L superior frontal cortex</td>
<td>8</td>
<td>-18</td>
<td>34</td>
<td>46</td>
<td>3.12</td>
</tr>
<tr>
<td>F temporal pole</td>
<td>21</td>
<td>48</td>
<td>8</td>
<td>-34</td>
<td>3.21</td>
</tr>
<tr>
<td>L temporal pole</td>
<td>21</td>
<td>-52</td>
<td>-10</td>
<td>-28</td>
<td>3.20</td>
</tr>
<tr>
<td>F insula</td>
<td>36</td>
<td>-14</td>
<td>16</td>
<td></td>
<td>3.34</td>
</tr>
<tr>
<td>L precentral gyrus</td>
<td>4</td>
<td>-22</td>
<td>-24</td>
<td>62</td>
<td>3.55</td>
</tr>
<tr>
<td>F superior temporal sulcus</td>
<td>21/37</td>
<td>52</td>
<td>-32</td>
<td>8</td>
<td>3.10</td>
</tr>
<tr>
<td>L superior temporal sulcus</td>
<td>21/37</td>
<td>-52</td>
<td>-40</td>
<td>10</td>
<td>3.42</td>
</tr>
<tr>
<td>(OI \textendash{} OI) \times (OE \textendash{} RE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No significant voxels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.4 Discussion

This study explored the neural systems involved in the observation and recognition of expressive and instrumental gestures. It was hypothesised that divergent neural pathways exist for the perception of these two categories of gesture. This was based on evidence that recognition of expressive gestures (but not instrumental) is impaired in children with autism (Attwood et al., 1988; Chapter 5). It is argued that understanding expressive gestures requires a concept of the mental states of others. It is thought therefore, that the failure to recognise these gestures in the autistic children is due to an inability to mentalise (Attwood et al., 1988), a characteristic deficit seen in individuals with autism (Frith et al., 1991). Consequently, it was predicted that the recognition of expressive gestures will involve brain regions previously associated with theory of mind tasks, in particular the anterior paracingulate cortex the right superior temporal sulcus and the temporal poles bilaterally (see Frith, 2001 for review).

In the current study an interaction demonstrated five regions of significant activity associated with recognising expressive gestures. These were the left medial prefrontal/anterior paracingulate cortex (BA 32/9), the temporal poles bilaterally and the right and left superior temporal sulci (BA 21/37). Two other regions that activated, but not significantly, were the right insula and the left precentral gyrus. A conjunction analysis demonstrated a number of different regions to be
associated with expressive gestures compared to instrumental gestures. These were the right medial orbito-frontal cortex (BA 11), the right and left amygdala, the right superior temporal sulcus (BA 21/37) and the right and left dorsal prestriate cortices (BA 17). The reverse of this contrast yielded activity in the left inferior frontal cortex (BA 44) and the left middle frontal cortex (BA 10) only. A further conjunction analysis of both the instrumental and expressive contrasts demonstrated regions of the brain associated with recognising gestures compared to observing gestures. These regions were the left inferior frontal cortex (BA 47 and 45), the left lateral orbito-frontal cortex (BA 11/47), the temporal poles bilaterally (BA 20), the middle temporal cortex bilaterally, the right superior temporal sulcus (BA 21) and the cerebellum. The reverse of this contrast yielded activity in the medial fronto-polar cortex bilaterally (BA 10), the right posterior cingulate cortex (BA 31) and the right angular gyrus (BA 39)

6.4.1 Expressive gestures

6.4.1.1 The superior temporal sulci

In accordance with the prediction the superior temporal sulcus was found to be involved in the processing of expressive gestures as demonstrated by the interaction analysis (see figure 6-6). Moreover, the right STS was found to be even more active during the recognition of the expressive gestures. This result is consistent with the study by Bonda et al. (1996) which also compared two types of biological motion (expressive and goal directed). In this study (Bonda et al., 1996) the volunteers were asked to pay attention to the overall pattern of
movement in order to undertake a recognition memory test at the end of each scanned condition. However, in the study by Bonda, (Bonda et al., 1996) activation of the right STS was only seen when the expressive body movement condition was compared with a random motion condition. This activation did not survive when expressive body movement was compared to the goal directed hand movement. However, the goal directed hand movement condition compared with random motion did not elicit any right STS activation. Taken together, the results of these two studies indicate that activity in the right STS is specific to interpreting and recognising expressive body motion.

A number of other studies have activated the superior temporal sulcus in relation to various types of biological motion including hand actions (Grezes et al., 1998; Grezes et al., 1999) body movements (Senior et al., 2000; Grossman et al., 2001) mouth movements and lip reading (Puce et al., 1998; Calvert et al., 1997) and eye movements or gaze-direction (Puce et al., 1998; Wicker et al., 1998; Hoffman and Haxby, 2000). However the specific role of this region remains unclear. A study by Narumoto et al., (2001) found that the right STS was part of a network of face-responsive brain regions. They found that selective attention to facial emotion specifically enhanced the activity of the right STS compared with attention to the face per se. They suggest that the right STS plays a specialised role in facial emotion recognition within a distributed face-processing system. However, in the current study the facial expressions of the actors were obscured. The expressive
gestures themselves conveyed emotional states and thus in this instance it is perhaps the attention to emotion that the right STS is sensitive too.

Recent studies suggest that implied biological motion also activates this region on the right (Senior et al., 2000). Kourtzi and Kanwisher, (2000) imaged volunteers while they viewed static images which implied motion (e.g. a discus thrower captured in the act of throwing) and static images that did not imply motion (e.g. the discus thrower at rest). They found stronger activation of the STS in association with implied motion. Taken together with the results of the neuroimaging studies of theory of mind (Fletcher et al., 1995; Goel et al., 1995; Chapter 3) and intention attribution (Brunet et al., 2000; Castelli et al., 2000) that have also activated this region, these findings suggest that the right STS is sensitive to stimuli that signal the actions of another individual or that right STS is...
involved in social perception from visual cues (Allison et al., 2000; Narumoto et al., 2001).

However, the study by Vogeley et al. (2001) found the right superior temporal sulcus to be specifically involved in taking the self-perspective. They argue that both common and differential neural mechanisms exist for taking the self-perspective and modeling the minds of others. They employed an adaptation of the same story comprehension task which has been used previously for imaging studies of theory of mind ability (Fletcher et al., 1995; Happé et al., 1996; Chapter 3). This involved the inclusion of self-perspective conditions in which the subjects had to ascribe adequate behaviour to themselves in given situations. The main finding was a disassociation between the medial prefrontal cortex that activated during the mindreading and the self-perspective comparisons and the right temporo-parietal junction that activated during the self-perspective comparison only. Although this is the first study to attempt to discriminate anatomically between these two functions, the results do not account for the growing body of evidence from functional imaging studies involving different modes of self monitoring, which suggests that modeling the mind of one's self is at least anatomically comparable to modeling the minds of others; visual self-recognition (Kircher et al., 2000), memory for autobiographical events (Maguire and Mummery, 1999; Maguire et al., 2000), monitoring one's own speech (McGuire et al., 1996a.), monitoring thoughts (McGuire et al., 1996b.), perception of tickling (Blakemore et al., 1998), perception of pain (Rainville et al., 1997),
reporting emotional responses (Lane et al., 1997) and monitoring intended response (Carter et al., 1998). All of these studies have found activity in the paracingulate cortex associated with self-monitoring.

The current study and the study by Bonda et al. (1996) are the only neuroimaging studies of biological motion which discriminate between expressive actions, which involve theory of mind mechanisms, and goal directed or instrumental actions. Both these studies indicate that the right STS is specifically involved in the processing of the former category, while the left STS is involved in the latter.

6.4.1.2 The medial prefrontal cortex (BA 32/9)

The interaction also demonstrated that as hypothesised the medial prefrontal cortex (anterior paracingulate cortex, BA32/9) was involved in the perception of expressive gestures. Like the right superior temporal cortex this region activated more when these gestures were explicitly recognised (See figure 6-7). This was confirmed by a simple main effects analysis comparing recognition of expressive with recognition of instrumental gestures which found activity in both the right and left paracingulate cortex (BA 32/9) (right: x=2, y=58, z=12 (z=3.63); left: x= -8, y=52, z=8, (z=3.07)). This is consistent with the notion that in order to recognise expressive gestures, one requires a theory of mind. As discussed previously, activity in the anterior paracingulate has been attributed to a theory of mind mechanism in a number of neuroimaging studies (Goel et al., 1995; Fletcher et al., 1995, Brunet et al., 2000, Castelli et al., 2000, Vogeley et al., 2001,
McCabe et al., 2001, Chapter 3, Chapter 4. McCabe et al. (2001) and the study described in chapter 4 demonstrated that “cold” mentalising, or in other words reading the mind of another in the absence of visual or auditory behavioural cues, is mediated by the anterior paracingulate cortex alone.

Figure 6-7. Sagittal and coronal views of activation in the anterior paracingulate cortex elicited by recognition of expressive gestures (interaction: \((RE - OE) \times (OI - RI)\). The data shown is from the group mapped onto a standard template.

6.4.1.3 The temporal poles and the amygdala

Finally the perception of expressive gestures involves the amygdala bilaterally and the temporal poles adjacent to the amygdala. The temporal poles show the same interaction as the right STS and the anterior paracingulate cortex, with greater activity for the recognition of expressive gestures. The function of the temporal poles in relation to mentalising tasks or social perception is unclear. The primate temporal poles are considered to be a higher-order visual cortical area
specialised for object and face recognition and memory (Nakamura and Kubota, 1996). In functional neuroimaging studies of humans the temporal poles have activated in association with recollection of familiar faces and scenes (Nakamura et al., 2000), and emotional memory retrieval (Dolan et al., 2000). It seems reasonable that the recognition of expressive gestures involves emotionally laden episodic memory. The same possibly applies to mentalising in accordance with the simulation theory of theory of mind, discussed in chapter 1, which suggests that in order to understand another’s behaviour we automatically imagine ourself to be in the other’s situation by drawing on similar experiences.

In contrast to the temporal poles, the amygdala is active during both observation and recognition of expressive gestures, suggesting an automatic processing of expressions of feeling states. A functional imaging study of neural responses during the evaluation of trustworthiness (Winston et al., 2002) found bilateral activation of the amygdala in association with faces deemed untrustworthy, regardless of whether the volunteer was making an explicit judgement of trustworthiness or of age. They suggest that the brain responds automatically to threat related stimuli, which is in keeping with the findings of earlier studies (Morris et al., 1998; Whalen et al., 1998; Strange et al., 2000; Vuilleumier et al., 2001).

It is interesting that the three areas implicated in previous studies of theory of mind (Frith, 2001); the anterior paracingulate, the right STS and the temporal
poles, are significantly more active during the recognition of expressive gestures, while the amygdala automatically activates independent of task. The Winston study (Winston et al., 2002) found a similar association between the right STS and the amygdala in that amygdala activation was automatic but the right STS showed enhanced signal change when the volunteers were required to make explicit judgements of trustworthiness.

There have been recent claims that the amygdala may be involved in the mediation of theory of mind (Baron-Cohen et al., 1999). An fMRI study by Baron-Cohen et al. (1999) reported activation in the left amygdala alone during a task in which normal volunteers were required to infer the mental state of an individual from the expression of their eyes. They also found significantly less amygdala activation when adults with high functioning autism and Asperger syndrome performed the same task. Fine et al. (2001) reported a single case study of a patient with early or congenital left amygdala damage. They found their patient to be severely impaired in his ability to represent mental states.

Other neuroimaging studies have demonstrated a role for the amygdala in processing facial expressions (Morris et al. 1996), monitoring gaze (Kawashima et al. 1999; Wicker et al. 1998) and the perception of biological motion (Bonda et al., 1996). Single cell recordings in the monkey amygdala have shown selective responses to facial expressions of threat, predator warning vocalisations, infants’ vocalisations on being separated from the mother and displays of a conspecific
approaching (Brothers and Ring, 1992; Kling and Brothers, 1992). Thus the amygdala appears to be involved in recognising emotional behaviour. This would account for the activation in the current study and the study by Baron-Cohen et al. (1999).

6.4.2 Instrumental gestures

6.4.2.1 Left inferior frontal cortex

Only two regions activated in association with the recognition of instrumental gestures compared to expressive. These were the left inferior frontal cortex (BA 44) and the left middle frontal cortex (BA 6/8) (see figure 6-8). This left lateralisation suggests that the instrumental gestures evolve from within a language-based network. In fact, language and gesture are closely linked developmentally leading to the suggestion that language is built from existing abilities such as attention, perception, imitation and symbolic processing, and mediated by common neural systems (Bates and Dick, 2002). Brodmann area 44, commonly known as Broca’s area (Tomaiuolo et al., 1999) – the motor area for speech production, has been associated with semantic processing of language (Shaywitz et al., 1995; Buckner et al., 1995; Gabrielli et al., 1996). However, a functional imaging study of stimulus and task induced semantic processing found this region to be activated by both repetition (silently mouthing a word) and semantic decision, suggesting a general role for this region in stimulus-driven semantic and phonological processing (Noppeney and Price, 2002). PET studies have shown that BA 44/45 is more active when people are passively watching
meaningful arm movements than when they are passively watching static hands, and that BA 44/45 is not more active during passive observation of meaningless arm movements (i.e. movements that cannot be named) (Grezes et al., 1998). This may account for the activation seen here as name retrieval of the recognised instrumental gestures takes place automatically in this task. Also, the developmental literature has shown that slightly before the onset of naming around 12 months, children start to reproduce brief actions associated with specific objects. A review of the literature by Shore et al. (1990) has shown that "recognitory gestures" and vocal naming emerge around the same time and are positively correlated across children in the period of 12 to 18 months, (i.e. early gesturers tend to be early namers).

Functional imaging research into the neural correlates of the perception of British sign language is consistent with the idea of left hemisphere processing of gestures relating to language processing (MacSweeney et al., 2002). In this study, the authors compared brain activity in hearing and congenitally deaf native users of British sign language and hearing, non-signing volunteers, during an audio visual sentence acceptability task. They found similar patterns of activity in all three groups in the inferior prefrontal regions bilaterally and superior temporal regions bilaterally, though predominantly on the left. An earlier study had suggested that sign language makes special demands on the right hemisphere (Neville et al., 1998). However, in the MacSweeney study (MacSweeney et al., 2002) which directly compared natural language inputs, this was not the case. In this study, the
stimuli predominantly involved sentences that did not express emotions, which makes them more comparable with the instrumental gestures of the current study. It might be interesting to directly compare the native signers' perception of ‘signed’ sentences which involve emotional expression such as I am happy, I don't care with ‘signed’ instrumental sentences, such as look over there, come to me, to see if the results corroborate with the current study.

There is strong evidence however, from an fMRI study that BA 44 is involved in imitation. Iacoboni et al., (1999) found more activity in BA 44 when volunteers lifted one of two fingers cued by a video of the corresponding finger movement, than when choice of finger movement was cued by a static hand with a dot on the to-be-moved finger, or by a dot in one of two positions on a blank rectangle. BA 44 is considered the human homologue of monkey area f5 which contains so-called “mirror neurons”, which are activated by observation of, and execution of grasping actions Rizzolati et al. (1999). The function of BA 44 in imitation remains unclear however it may be involved in translating perceptual input from observation of a model’s movements into matching motor output. Alternatively, it has been speculated that this region is activated by imitation only on the occasions when imitation is accompanied or mediated by movement recognition (for review see Heyes, 2001).
Figure 6-8. Sagittal and coronal views of activation in the left inferior frontal gyrus and middle frontal gyrus elicited by recognition of instrumental gestures compared to expressive gestures (conjunction). The data shown is from the group mapped onto a standard template.

6.5 Conclusions

The results of this study indicate that as predicted, dissociable neural networks exist for the perception of instrumental and expressive gestures. The perception of expressive gestures involves the anterior paracingulate cortex, the amygdala, the temporal poles bilaterally and the right superior temporal sulcus. These regions have all previously been implicated in a neural network constituting a "social brain". In contrast, instrumental gestures involve a left-lateralised system correlated with semantic and phonological language areas.
Three regions in particular are more active during the recognition of expressive gestures. These are the anterior paracingulate cortex, the right STS and the temporal poles bilaterally. These three regions have consistently activated in previous functional imaging studies of theory of mind ability, over and above what is expected from the specific demands of the tasks. This recurring pattern of increased activity in the anterior paracingulate cortex, right STS, and the temporal poles, in association with mentalising or tasks involving mentalising suggests that these three regions form part of a network involved in processing the intentions of biological agents (Frith 2001).

The involvement of the anterior paracingulate cortex in the perception of expressive gestures indicates that a theory of mind is indeed necessary for the recognition of these gestures and accounts for the marked impairment on this task demonstrated by autistic individuals.
7 Chapter 7: Discussion

7.1 Summary of the studies described in this thesis

"Theory of mind" or mentalising is the ability to attribute independent mental states to self and others in order to explain and predict behaviour. It is a highly sophisticated and ubiquitous element of social cognition. This thesis has described a series of experiments which examine neural and behavioural characteristics of "theory of mind".

Chapter 3 described an fMRI study that sought to examine anatomical convergence between mentalising tasks presented in different domains; verbal (story comprehension) and non-verbal (understanding the joke in captionless cartoons). Each modality consisted of three conditions: 1) theory of mind, in which one or more characters involved held a false belief, 2) non-theory of mind, in which a physical cause could be ascribed to explain a situation, and 3) a control condition, comprising either unlinked sentences or jumbled pictures. A conjunction analysis of the verbal and non-verbal mentalising conditions showed considerable overlap in brain activation, specifically in the medial prefrontal cortex which was pinpointed to be the anterior paracingulate cortex (x=-10, y=48, z=12). This was the only region of activity found to be uniquely associated with mentalising. Activation was also seen in the temporo-parietal junctions, particularly on the right, and the temporal poles bilaterally. However, these
regions were also activated during the non-theory of mind tasks, only to a much lesser extent. Thus, they were not attributed to mentalising *per se*. In this study the volunteers were required to consider a scenario and retrospectively explain the behaviour of the person or persons involved. Theory of mind ability is an automatic process that allows us to both explain *and* predict the behaviour of others. One potential criticism, then, of this study is the assumption that "off-line" tasks reflect mentalising in real-time, whereas in practice this may potentially involve different mechanisms.

The PET experiment described in Chapter 4 attempted to address this issue by employing an "on-line" paradigm that also allowed tight control of other cognitive demands. The main aim of this study was to examine whether the same neural regions are activated during off- and on-line mentalising. A secondary aim of this study was to isolate a particular component of mentalising; the adoption of an intentional stance (Dennett, 1971; Dennett, 1991; Dennett, 1996). This is the disposition to treat an entity as a rational agent, possessing particular beliefs, desires and intentions. In this study volunteers played a computerised version of the children's game 'stone, paper, scissors'. In the mentalising condition volunteers believed they were playing against the experimenter and thus adopted an "intentional stance". In the comparison condition, volunteers believed they were playing against a computer which used pre-programmed rules to generate a response, and thus treated their opponent as a non-rational entity. In fact, in both instances during the critical scan window they played against a random sequence.
The only difference between the conditions was the attitude, or 'stance', adopted by the volunteer. In a third condition the volunteers were told they were playing against a random sequence. It was inferred that the rule solving condition engaged the same higher cognitive processes as the mentalising condition with the exception of the “mentalising” component. The random generation condition involved fewer cognitive processes, although these were all engaged by the rule solving and the mentalising conditions. Thus the random generation condition acted as a low-level control. The main comparison of the mentalising condition versus rule solving showed only one region of significant activation; the anterior paracingulate cortex bilaterally (x=8, y=54, z=12). No further regions appeared even when the statistical threshold was lowered to p=0.1. This is consistent with the results from chapter 3 and previous studies of mentalising (Fletcher et al., 1995; Goel et al., 1995; Baron-Cohen et al., 1999; Brunet et al., 2000; Castellie et al., 2000; Vogeley et al., 2001; McCabe et al., 2001). Although unlike these studies which have also activated a range of additional regions of the brain, this was the only region to be activated in association with mentalising. In this study none of the conditions involved any visible behaviour from the opponent such as facial expression, eye gaze or body posture, which might have elicited supplementary activity in a more general neural network of social cognition. Were the volunteers to play this game face to face with their opponent a more widespread network of activity involved in processing these additional social cues might be expected, such as the right temporo-parietal junction and the temporal poles which activated in chapter 3.
Chapter 7 Discussion

The experiment described in chapter 5 looked at the contribution of mentalising skills to everyday social behaviour. This study contrasted four populations of children; normal 4-, 5-, and 6-year olds and children with autism, on their ability to recognise two different categories of communicative gesture; instrumental and expressive. The latter are gestures which express feelings and beliefs (e.g. I’m angry or “I’m happy”). The recognition of which is considered to require a concept of mental states. On the other hand instrumental gestures communicate commands or requests and aim to influence the immediate behaviour of others (e.g. “come here”, or “look over there”). The aims of the study were to ascertain the age at which children developed an understanding of gesture and expose differences which may indicate the existence of dissociable neural networks for the perception of these two categories of gesture. The children were shown short videos of the two types of gesture which were presented in a random order and asked for the meaning of each gesture. The performance of the normal groups of children demonstrated that by 4-years of age a child has an understanding of both instrumental and expressive gestures. This performance continues to improve through to the age of 6 ½-years, by which time it is virtually at ceiling on the more common gestures studied here. The greatest improvement was seen between the ages of 4- and 5-years of age. These results also showed no differences in developmental age between instrumental and expressive gestures. However, the children with autism demonstrated highly significant differences in their performance on instrumental and expressive gestures, displaying a marked
impairment on recognition of the latter type. This was attributed to the characteristic deficit in theory of mind ability displayed in these children (Frith et al. 1991). This result is evidence for dissociable neural networks underpinning expressive and instrumental gestures.

An fMRI study to determine the neural correlates of these two potential networks, in normal adults, was described in chapter 6. The main aim was to compare the observation and recognition of instrumental with that of expressive gestures. Volunteers were shown short videos depicting a range of gestures and were required to make a judgement about either the meaning (observe and recognise) or physical characteristics (observe). It was predicted that the perception of expressive gestures would involve a network of brain regions previously found to be associated with theory of mind. The results demonstrated that the observation and recognition of expressive gestures does indeed engage different neural regions to the observation and recognition of instrumental gestures. Recognition of expressive gestures activated the anterior paracinguate cortex, the right superior temporal sulcus and the temporal poles bilaterally. The amygdala also activated in response to expressive gestures, regardless of the task. Whereas instrumental gestures were associated with activity in a left lateralised language based circuit, the inferior frontal and middle frontal cortex. All the regions activated by expressive gestures either activated in previous functional imaging studies of theory of mind, including chapter 3 or have been linked with a circuit for social intelligence (Baron-Cohen et al., 1999). This suggests that recognition
of expressive gestures does involve a specialised neural circuit of theory of mind and social perception and accounts for the impaired performance of the children with autism on the understanding of these gestures.

7-1-1 Summary of the main findings of this thesis

The experiments described in this thesis have examined the neural correlates of theory of mind and a broader framework of social cognition, by testing performance on both mentalising tasks or tasks involving mentalising skills, with both functional neuroimaging and behavioural methods. The main findings are summarised here. Firstly, expressive and instrumental gestures have divergent neural pathways that can be selectively damaged resulting in an impaired ability to recognise such gestures. Recognition of expressive gestures is thought to require a theory of mind. In accordance with this idea chapter 6 has demonstrated that such processing engages a network of neural regions that includes the anterior paracingulate cortex, the right superior temporal sulcus, the temporal poles and the amygdala. These neural regions have all previously been implicated in a model of a “social brain” (Baron-Cohen et al., 1999). The precise functions of these regions within this model have yet to be established. However, the studies in this thesis have provided compelling evidence that a circumscribed region of the anterior paracingulate cortex (BA 32/9) mediates our ability to mentalise and is independent of modality (either verbal or non-verbal) or whether the stimulus is retrospective or on-line. This supports the notion of an innate, dedicated and domain-specific cognitive mechanism thought to underpin this ability (Leslie,
1987). The other regions of the social brain; the amygdala and temporal poles, orbito-frontal cortex and the right superior temporal sulcus may form part of a network involved in processing explicit social behaviour such as facial expression, gaze direction, detection of the behaviour of agents and the analysis of goals and outcomes (Frith and Frith, 1999). The next section discusses each of these brain regions in more detail and speculates further on their roles within social cognition.

7.2 The “Social Brain”

7.2.1 The anterior paracingulate cortex

The results of this thesis have demonstrated that a circumscribed region of the anterior paracingulate cortex (BA 32/8/9) is activated by mentalising tasks (Chapters 3 and 4) or by tasks involving mentalising abilities (Chapter 6) (See figure 7-1). This main finding is consistent with the results of previous neuroimaging studies which have also demonstrated activity in the medial prefrontal cortex/anterior paracingulate cortex in association with mentalising tasks (Fletcher et al., 1995; Goel et al., 1995; Baron-Cohen et al., 1999; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 20001; McCabe et al., 20001)(see figure 7-2). This region can be described as the most anterior part of the paracingulate cortex, where it lies anterior to the genu of the corpus callosum and the anterior cingulate cortex. The paracingulate cortex (BA 32) is often considered to be part of the anterior cingulate cortex (ACC) that incorporates the
cytoarchitectural areas of BA 24, 25 and 33. The ACC is an ancient structure which has been broadly defined by Broca as belonging to the limbic lobe (Bush et al., 2000). However the existence of an unusual type of projection neuron (spindle cell) found in the subareas of the ACC 24a, 24b, and 24c in the human, and some other higher primates (pongids and hominids) is evidence that the ACC has undergone changes in recent evolution (Nimchinsky et al., 1999). The ACC occupies two tiers of the medial wall of the cerebral hemisphere. The ventral or limbic tier occupies the surface of the cingulate gyrus and incorporates BA 24a, 24b and 25. The dorsal or paralimbic tier is deeply buried in the cingulate sulcus and incorporates BA 24c, 32 and 33. In the human brain BA 32 often extends anteriorly into the paracingulate gyrus (Paus et al., 2001). This is dependent on the presence or absence of a paracincingulate sulcus the incidence of which is aproximately 50% (Paus et al., 2001). The presence of a paracingulate sulcus indicates that BA 32 is likely to be located on a gyral crown and this increased cortical folding may be indicative of a progressive evolution of this region in humans (Zilles et al., 1988). BA 32 has been described as cytoarchitectonically a cingulo-frontal transition area (Devinsky et al., 1995) and therefore anatomically (and speculatively functionally) distinct from the anterior cingulate cortex proper. As theory of mind is almost entirely unique to humans it is reasonable to hypothesise that a mechanism underpinning this ability will be sited in the neocortex (prefrontal cortex) or an area of the brain that has undergone recent evolutionary changes.
Figure 7-1

Points of maximum activity in the anterior paracingulate cortex associated with mentalising from the functional imaging studies described in this thesis. Activations are mapped onto a. single subject T1 weighted MR image template and b. glass brain.

Figure 7-1a

Figure 7-1b

- “Theory of mind” stories and cartoons
- Adopting an intentional stance
- Expressive gestures
Figure 7-2
Points of maximum activity in the anterior paracingulate cortex associated with mentalising from previous published studies of mentalising, mapped onto figure 7-1b.

a - Fletcher et al., 1995
b - Goel et al., 1995
c - Baron-Cohen et al., 1999
d - Brunet et al., 2000
e - Castelli et al., 2000
f - Vogeley et al., 2001
g - McCabe et al., 2001

Chapters 3 and 4 have shown that theory of mind ability is mediated by this circumscribed region. However, there is, as yet, no computational account of the cognitive mechanism that underlies this ability and therefore we cannot assign a precise role to the paracingulate cortex. McCabe et al. (2001) suggest that this region acts as an active convergence zone which, during their cooperation task, "binds joint attention to mutual gains with sufficient inhibition of immediate reward gratification to allow cooperative decisions". However, this account is specific to their task and does not apply to the competitive task described in chapter 4 or to the other mentalising tasks that have activated this region.
The findings of chapter 4 in which the anterior paracingulate cortex was the only region activated by the "cold" mentalising task suggest that this region alone is responsible for the theory of mind mechanism. This is consistent with Leslie's model of mentalising (Leslie, 1987; 1994; Baron Cohen, 1995) which proposes an innate and domain specific mechanism. Leslie (1987; 1994) suggests that the theory of mind mechanism (ToMM) depends upon a representation of imaginary circumstances "decoupled" from reality. When explaining a person's behaviour in terms of a belief we have to recognize that this belief might not correspond to reality. Even when it conflicts with reality it is the belief not the reality that determines behaviour. We also have to distinguish the representation of the reality by the other person from our representation of reality. Perhaps activity in paracingulate cortex is necessary to maintain these decoupled representations in the face of competition from representations of the actual state of the world. This function is analogous to the function of the adjacent, more posterior part of ACC. This region is active in tasks which elicit response conflict (Barch et al., 2001). In the Stroop task a series of colour words (e.g. the word RED written in green ink) are presented, and the volunteer has to name the colour of the ink. This less practiced response conflicts with the more usual response of reading the word. One may speculate that paracingulate cortex has a similar role in handling possible conflict between actual and virtual representations, but in the more abstract domain of mental states rather than motor responses. This ability to
handle different representations of the world is also an essential requirement for adopting an intentional stance.

Nonetheless, this region appears to be strongly associated with a more widespread network of brain regions involved in social cognition. Frith and Frith (1999) have suggested that the ability to make inferences about other people’s mental states has evolved from pre-existing abilities. They suggest abilities that are relevant to mentalising should include 1) the ability to distinguish between animate and inanimate entities, 2) the ability to share attention by following the gaze of another agent, 3) the ability to represent goal-directed actions and 4) the ability to distinguish between the actions of the self and others (Frith and Frith, 1999). It is possibly these abilities and functions that account for the activity seen in the other neural regions which have activated in association with mentalising in this thesis, in particular the right superior temporal sulcus and the temporal poles which will be discussed in the next sections.

7.2.2 The right superior temporal sulcus

The right superior temporal sulcus (STS) has activated in a number of studies in association with theory of mind ability. In chapter 3 both right and left STS were associated with understanding the meaning of stories and cartoons involving people, with or without the requirement to mentalise. In chapter 6, right STS activity was associated with the recognition of communicative hand gestures which express feelings and mental states. The previous functional neuroimaging
literature of theory of mind tasks has associated STS activity with judging mental states from the expressions of another’s eyes (Baron-Cohen et al., 1999), the involvement of people when understanding causality and intentionality (Brunet et al., 2000), the attribution of intentions to the movements of geometric shapes (Castelli et al., 2000) and taking the self perspective (Vogeley et al., 2001). It is likely that all of these tasks share one common function which is able to account for this consistent activation.

Functional imaging studies which examine the perception of biological motion activate the right STS in association with hand actions, (Grezes et al., 1998; Grezes et al., 1999) body movements (Bonda et al., 1996; Senior et al., 2000; Grossman et al., 2001) mouth movements and lip reading (Puce et al., 1998; Calvert et al., 1997), implied biological motion (Senior et al., 2000; Kourtzi and Kanwisher, 2000) and eye movements or gaze-direction (Puce et al., 1998; Wicker et al., 1998; Hoffman and Haxby, 2000; Calder et al., 2002). Also, Narumoto et al. (2001) found that the right STS was part of a network of face-responsive brain regions. Selective attention to facial emotion specifically enhanced the activity of the right STS compared with attention to the face per se. They attribute this activation to facial emotion recognition within a distributed face-processing system.

Consideration of all of the above findings has led researchers to speculate on the role of this region within a social cognitive network. Allison et al. (2000) suggest
that the right STS is sensitive to stimuli that signal the actions and intentions of another individual. The results of neuroimaging, electrophysiological and single cell recording studies converge to suggest that initial analysis of social cues occurs in the STS region, which is anatomically well sited to integrate information derived from both the ventral and dorsal visual pathways. In accordance with this notion Frith and Frith (1999) suggest that the STS is involved in the detection of the behaviour of agents and analysis of the goals and outcomes of this behaviour. The data from this thesis fits this theory well. The results of the experiment in chapter 3 in particular, indicate that the right STS is involved in explaining the behaviour of others as a result of recognising both a physical cause or their mental states. Frith and Frith’s idea (Frith and Frith, 1999) gains support from a recent fMRI study of social judgement. Winston et al. (2002) asked subjects to view faces and make judgements on either trustworthiness or age. They found right STS activity when volunteers made explicit judgements about trustworthiness. The authors attribute this activity to intention detection from visual cues, a critical component in determining whether or not to trust an individual.

7.2.3 The temporal poles

The final regions of the brain that consistently activate during mentalising tasks are the temporal poles bilaterally. The temporal poles are generally associated with object and face recognition in primates (Nakamura and Kubota, 1996). However, functional imaging studies of humans have activated the temporal poles
in association with a broader context of episodic memory retrieval in visual and auditory domains. Thus the temporal poles are active during the recollection of familiar faces and scenes (Nakamura et al., 2000), the recognition of familiar voices (Nakamura et al., 2001) emotional memory retrieval (Dolan et al., 2000) and autobiographical memory retrieval (Fink et al., 1996). Taken together these results suggest that the temporal poles are a store for personal semantic and episodic memories. As discussed in chapter 6, the relationship between this and mentalising may lie in the simulation theory of theory of mind. By trying to explain or predict another’s behaviour we implicitly imagine ourselves in their situation and thus draw upon and project our own experiences onto that person. It is thus reasonable to suggest that this particular function of the temporal poles is an essential prerequisite for acquisition of a theory of mind and may go some way to explain the time course of the development of this ability. A child’s experiences increase with time providing him with a broader range of “schemas” to refer to when attempting to explain the behaviour of other people, which will improve the probability of the child’s accuracy at this task. According to Leslie (Leslie, 1987) the theory of mind ability is innate, but it is believed to have evolved from pre-existing abilities such as the perception of eye gaze, movements of conspecifics and the ability to imitate (Frith, 2001). These abilities are thought to extract information about the social environment which can be interpreted by the theory of mind mechanism. The more information provided the more the ability is refined.
The voxel-based morphometric study by Abell et al. (1999) found significant increases in the middle and inferior temporal cortices and the peri-amygdaloid cortex in patients with Asperger syndrome. The same study found further abnormalities in the right STS and anterior paracingulate. These results suggest that this system of brain regions has failed to develop normally in the Asperger group. Taking together the findings of previous functional imaging studies of theory of mind ability (Fletcher et al., 1995; Goel et al., 1995; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001) these results suggest that all these regions play an important role in the early development and acquisition of a normal theory of mind.

7.2.4 Associated neural regions

The “social brain” as proposed by Brothers (Brothers 1990; Brothers and Ring, 1992) and extended by Baron-Cohen (Baron-Cohen et al., 1999; 2000) incorporates two other regions; the orbito-frontal cortex and the amygdala. A role for the amygdala is consistent with the results of chapter 6 which found activity bilaterally in this region when the perception of expressive gestures was compared with that of instrumental. There have also been suggestions that the amygdala may play a part in either the development, or functioning, of theory of mind (Baron-Cohen et al., 1999; Baron-Cohen et al., 1999; Fine et al., 2001). Activity during the perception of expressive gestures is correlated with theory of mind, as recognition of these gestures is thought to require mentalising skills. However, among all the functional imaging studies of theory of mind (Baron-
Cohen et al., 1994; Fletcher et al., 1995; Goel et al., 1995; Baron-Cohen et al., 1999; Chapter 3; Brunet et al., 2000; Castelli et al., 2000; Vogeley et al., 2001; McCabe et al., 2001), only Baron-Cohen et al. (1999) found amygdala activation. In this study volunteers were required to read a mental or emotional state from an individual’s eyes. It is feasible that this finding may reflect the amygdala’s responsiveness to eye gaze information (Kawashima et al., 1999) or, as discussed in chapter 6, the recognition of emotional behaviour, which would account for its involvement during the observation and recognition of expressive gestures.

The results of chapter 6 demonstrated that the amygdala activates automatically when viewing expressive gestures, regardless of whether the processing is implicit or explicit. This was in contrast to the activity in the anterior paracingulate, right STS and temporal poles which showed significant increases when volunteers explicitly recognised these gestures. This result is consistent with an earlier study of social cognition (Winston et al., 2002) which found automatic amygdala activation in response to untrustworthy faces, independent of whether the volunteer was making judgements about gender or explicitly judging trustworthiness. This implicit response of the amygdala may reflect early processing of socially salient information, essential for developing a theory of mind. One may speculate that the response of the amygdala becomes automatic as a child’s theory of mind ability develops, thus this region may be more active in younger children who have yet to fully develop a theory of mind. If so, one would predict that as a child’s theory of mind ability matures, the amygdala is recruited.
less and less. This would explain the theory of mind impairments seen in a patient who suffered congenital lesions of the amygdala bilaterally (Fine et al., 2001). The findings of this thesis suggest that while the amygdala may yet prove to be an essential pre-requisite to the appropriate development of theory of mind (Fine et al., 2001), it is perhaps not involved in the on-line representation of the mental states of others.

Baron-Cohen and Ring (1994) have proposed that orbito-frontal cortex is also part of a theory of mind circuit and suggested that damage to this region would produce subtle impairments in theory of mind, rather than the complete loss of mentalising ability. However, like the amygdala only one neuroimaging study of theory of mind to date has found increased activity in the orbito-frontal cortex (Baron-Cohen et al., 1999) and this was relative to the left fronto-polar region. No other regions were analysed in this study and so interpretation of the results is limited.

Clinical observations in humans and experimental reports in primates have consistently indicated that the orbito-frontal cortex is engaged in the regulation of social behaviour (Damasio and Van Hoesen, 1983; Stone, 2000). A recent study by Stone et al., (1998) found that patients with bilateral damage to orbito-frontal cortex were found to be impaired on a sophisticated test of theory of mind – the faux pas test, leading to suggestions that this area of the brain is important to theory of mind. However, these patients, who did not suffer any damage to the
paracingulate region of the brain, performed well on standard first and second-order false belief tasks. This indicates that their theory of mind abilities remained intact. Recognising a faux pas requires both an understanding of false or mistaken belief and an empathic inference of the affect it has on someone. In a recent neuroimaging study examining the neural basis of social norm violations (Berthoz et al. submitted) the orbito-frontal cortex activated in association with both intentional and unintentional (embarrassing) violations of social norms. They suggest that orbito-frontal cortex is a component of a system that responds to the aversive reactions of others, in particular others' anger. This may account for the orbito-frontal activation (non-significant) seen in chapter 6 as many of the expressive gestures were of an aversive kind.

However, it has been argued that the orbito-frontal cortex may play a similar role to the anterior paracingulate cortex in the mediation of theory of mind (Stone et al., 2000). The results of this thesis have found no evidence to support this as the orbito-frontal cortex has failed to activate in any of the studies directly examining theory of mind (Chapters 3 and 4). Previous neuroimaging studies of theory of mind which have activated the anterior paracingulate cortex but not the orbito-frontal cortex, have largely tended to test the ability to make inferences about cognitive states such as beliefs, knowledge, goals etc. (Goel et al., 1995; Fletcher et al., 1995, Chapters 3 and 4; Brunet et al., 2000, Castelli et al., 2000, Vogeley et al., 2001). Thus, this argument proposes that the orbito-frontal cortex is responsible for assessing affective or emotional theory of mind and the anterior
paracingulate region of the brain, performed well on standard first and second-order false belief tasks. This indicates that their theory of mind abilities remained intact. Recognising a faux pas requires both an understanding of false or mistaken belief and an empathic inference of the affect it has on someone. In a recent neuroimaging study examining the neural basis of social norm violations (Berthoz et al. submitted) the orbito-frontal cortex activated in association with both intentional and unintentional (embarrassing) violations of social norms. They suggest that orbito-frontal cortex is a component of a system that responds to the aversive reactions of others, in particular others’ anger. This may account for the orbito-frontal activation (non-significant) seen in chapter 6 as many of the expressive gestures were of an aversive kind.

However, it has been argued that the orbito-frontal cortex may play a similar role to the anterior paracingulate cortex in the mediation of theory of mind (Stone et al., 2000). The results of this thesis have found no evidence to support this as the orbito-frontal cortex has failed to activate in any of the studies directly examining theory of mind (Chapters 3 and 4). Previous neuroimaging studies of theory of mind which have activated the anterior paracingulate cortex but not the orbito-frontal cortex, have largely tended to test the ability to make inferences about cognitive states such as beliefs, knowledge, goals etc. (Goel et al., 1995; Fletcher et al., 1995, Chapters 3 and 4; Brunet et al., 2000, Castelli et al., 2000, Vogeley et al., 2001). Thus, this argument proposes that the orbito-frontal cortex is responsible for assessing affective or emotional theory of mind and the anterior
paracingulate is responsible for more cognitive metarepresentation. In order to test this theory a new experimental design is required, which can compare an affective theory of mind task with a cognitive theory of mind task. Unfortunately, none of the studies in this thesis are so designed, however this is a direction in which theory of mind research needs to travel. Nonetheless, the notion of dissociable psychological and neural mechanisms for cognitive and affective mental representations violates Leslie’s model of the theory of mind mechanism (Leslie, 1987) which has been supported by the findings of this thesis. This proposes an innate, domain specific neural mechanism for the “decoupling” of real and imagined representations, or in other words metarepresentations. It seems unlikely for the brain to have evolved two specialised regions for such a unique cognitive ability. The fact that orbito-frontal cortex is only associated with theory of mind when affective, particularly aversive stimuli are involved may reflect analysis of specific information which contributes to the outcome of the mentalising process. Therefore, while the orbito-frontal cortex appears to form part of the social brain with respect to processing aversive, or more generally affective stimuli, it is unlikely to be directly responsible for the mediation of theory of mind.
Chapter 7 Discussion

7.3 The neural correlates of autism

It was suggested earlier that the investigation of the neural correlates of theory of mind is driven by the hope of a better understanding of the biological basis of autism. It was discussed earlier, in chapter 1, that the deficit in theory of mind ability shown by autistic individuals probably results from a malfunction early on in the development of either the brain regions directly involved in mentalising or auxiliary brain networks/regions required for appropriate development of this capacity.

The findings of this thesis have highlighted three regions in particular that play an important role in theory of mind ability; the anterior paracingulate cortex, the right superior temporal sulcus and the temporal poles. All three of these regions were activated in chapter 6 in association with the recognition of expressive gestures, a function revealed by chapter 5 to be impaired in autistic individuals. None of the studies in this thesis and only a few previous functional imaging studies have directly examined the neural correlates of this ability in autistic individuals (Happé et al., 1995; Baron-Cohen et al. (1999); Castelli et al. (in press)). The previous studies (Happé et al., 1995; Baron-Cohen et al. (1999); Castelli et al. (in press)) have demonstrated reduced activity in either frontal regions or the amygdala in the autistic participants. In particular, Castelli et al. (in press), using the same intention attribution paradigm described earlier (Castelli et
al., 2000), demonstrated reduced activity in the anterior paracingulate, STS and temporal poles in a group of autistic adults compared to normal controls. The same study also demonstrated activity in the occipital gyrus comparable for both autistic and normal volunteers. However, functional connectivity showed that connectivity between this region and the temporo-parietal regions was weaker in the autistic individuals. The authors suggest that the weaker connectivity between the visual cortex and the STS reflects a top-down failure of feedback signals reaching the STS from the more anterior components of the "mentalising system".

The results of the experiment described in Chapter 4 suggest that the theory of mind mechanism is sited in the anterior paracingulate cortex alone. However, the consistent activation of the right STS during studies of explicit mentalising indicate that this region plays a significant role in the development of this ability, speculatively by providing early analysis of the goals and outcomes of explicit behaviour. This notion is supported by evidence that theory of mind deficits may be "acquired" by patients with right hemisphere stroke (Happé et al., 1999), an effect which was not demonstrated in matched left hemisphere stroke patients. This result suggests that healthy connections between these two areas are crucial for both the development and the everyday function of this ability. Indeed a voxel based morphometric study of individuals diagnosed with Asperger syndrome (Abell et al., 1999) found decreased grey matter in both these regions when compared to age and IQ matched controls.
A number of other regions were shown to be abnormal in the study by Abell et al. (1999) either as a result of decreased volume (left inferior frontal gyrus) or increased volume (amygdala/peri-amygdaloid cortex, middle and temporal poles and regions of the cerebellum). This findings suggests that these brain areas are part of a distributed network which in autism has failed to develop appropriately. This network is amygdala centered and there is much speculation about the role of the amygdala in autism (Baron-Cohen et al., 2000) and theory of mind ability. Unfortunately, the experiments in this thesis, with the exception of chapter 6, have failed to demonstrate amygdala activity in association with theory of mind. In chapter 6, amygdala activity was associated with the perception of expressive gestures but as discussed in the previous section this has been attributed to early processing of emotion. However, the experiments in this thesis were designed to examine the neural correlates of theory of mind ability and therefore the results do not speak directly to the biological basis of autism.

7.4 Future Perspectives

Some methodological issues could perhaps be addressed in future functional imaging studies of theory of mind. For instance, it would be interesting to record any physiological changes that would indicate increased levels of arousal occurring during any of the experimental conditions. A number of methods could be used for this that have been tried and tested in both PET and fMRI, such as skin conductance or pupil dilation measures (Critchley et al., 2000; Critchley et al., 2001).
Also, a recent MRI study has examined morphological variance of the anterior cingulate and paracingulate cortex (Yucel et al., 2001) in the left and right cerebral hemispheres. Three types of anterior cingulate patterns were identified determined by the absence or presence of a paracingulate sulcus; 1) prominent 2) present and 3) absent. The results demonstrated significant hemispheric differences across the whole sample. The paracingulate sulcus was found to be more commonly ‘prominent’ in the left hemisphere and more commonly absent in the right. They also demonstrated a significant gender difference in which males show an asymmetric pattern of fissurisation between hemispheres whereas females showed greater symmetry. These findings have implications on the interpretation of functional imaging data. Usually, group functional imaging data is smoothed and normalised to increase between subject averaging and facilitate reporting of the data in a stereotactic reference frame. Given the broad variation in gross morphology in this area it may be more appropriate to map single subject data from theory of mind studies directly onto high resolution images of the volunteers own brain, which will give a much more accurate localisation of the individual activations. Alternatively, the data could be mapped onto a flattened map of the cortex using a method called ‘cortical flattening’ which involves unfolding the cortical surface (Van Essen and Drury, 1997), this may lead to a dispersion of sites of activation in the resulting two dimensional map.
These structural variations by gender and hemisphere may also be of interest functionally. Future work could look for possible correlations between theory of mind ability using verbal and non-verbal tasks and structural and functional anatomical differences related to gender and possibly age. It would also be interesting to study the neuronal integrity of the anterior paracingulate cortex with MR spectroscopy in individuals with normal functioning and impaired ‘theory of mind’ in both autism and schizophrenia. Frith (1992) has proposed that schizophrenic individuals who show behavioural signs such as avolition, poverty of speech, social withdrawal, or incoherent or inappropriate speech or paranoid symptoms such as persecutory delusions, delusions of reference, or hearing other people’s voices are impaired on theory of mind tasks. In particular, paranoid patients are believed to over infer the mental states of others. The schizophrenic patients will make an interesting comparison for the autistic individuals as in autism it is thought that the ‘theory of mind’ mechanism, or as the results of this thesis suggest, the anterior paracingulate cortex, has failed to develop appropriately. However, most schizophrenic patients have experienced normal development of theory of mind ability which appears to be impaired following the onset of the disease.

The experiments in this thesis have added to our understanding of the neural substrates underpinning our ability to mentalise, the developmental characteristics of this ability and indirectly to the biological basis of autism which is thought to result from a breakdown in the early development of this ability. As well as
playing a significant role in answering these questions, the experiments also generate many more questions, giving scope for new experiments and theories about theory of mind. Neuroimaging is set to play a significant role in these questions, particularly in determining the precise functions of the neural substrates comprising the “social brain” and the mechanisms underlying theory of mind. The specific involvement of the amygdala in mentalising needs to be examined. Is the amygdala an essential prerequisite for acquiring a theory of mind? Or is its role to provide additional information for processing more heterogeneous social tasks such as faux pas? Imaging normal children, with fMRI, at different stages of social development may answer some of the enduring questions about the brain systems involved in the acquisition of this highly specialised ability - this may go some way to elucidating the brain basis of autism.
References


References


References


Buchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* 1997;7:768-78.


References


References


References


References


Lane RD, Fink GR, Chau PM, Dolan RJ. Neural activation during selective attention to subjective emotional responses. *Neuroreport* 1997;8:3969-72.


References


References


References


References


References

Raichle ME. Visualising the mind. Scientific America 1994;279:36-42.


References


Stuss DT, Gallup GG, Alexander M. The frontal lobes are necessary for "theory of mind". *Brain* 2001;**124**:279-86.


References


