Do ‘advanced’ theory of mind tasks offer a useful measure of mentalising ability in children with autism who have verbal and intellectual abilities within the normal range?

Ella Brent

D. Clin. Psy. 1999
University College London
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
<td>6</td>
</tr>
<tr>
<td><strong>Acknowledgements</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>Chapter One: Introduction</strong></td>
<td>9</td>
</tr>
<tr>
<td>1 Overview</td>
<td>9</td>
</tr>
<tr>
<td>2 Autism</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Historical Context</td>
<td>12</td>
</tr>
<tr>
<td>2.2 Definition and Diagnosis</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Epidemiology</td>
<td>17</td>
</tr>
<tr>
<td>2.3.1 Prevalence</td>
<td>17</td>
</tr>
<tr>
<td>2.3.2 Gender</td>
<td>17</td>
</tr>
<tr>
<td>2.3.3 Intellectual Impairment</td>
<td>17</td>
</tr>
<tr>
<td>2.3.4 Comorbidity</td>
<td>18</td>
</tr>
<tr>
<td>2.3.5 Social Class</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Course and Prognosis</td>
<td>19</td>
</tr>
<tr>
<td>2.5 Differential Diagnoses</td>
<td>19</td>
</tr>
<tr>
<td>2.5.1 Schizophrenia</td>
<td>20</td>
</tr>
<tr>
<td>2.5.2 Asperger syndrome</td>
<td>20</td>
</tr>
<tr>
<td>2.5.3 Semantic Pragmatic Disorder</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Causal Theories</td>
<td>22</td>
</tr>
<tr>
<td>3 Theory of Mind Hypothesis</td>
<td>25</td>
</tr>
<tr>
<td>3.1 False Beliefs</td>
<td>26</td>
</tr>
<tr>
<td>3.2 Radical Shift or Matter of Degree?</td>
<td>28</td>
</tr>
<tr>
<td>4 Development of Theory of Mind in Autism</td>
<td>30</td>
</tr>
<tr>
<td>5 Criticisms of the Theory of Mind Account</td>
<td>33</td>
</tr>
<tr>
<td>5.1 Imaginative impairments</td>
<td>33</td>
</tr>
<tr>
<td>5.2 Language impairments</td>
<td>34</td>
</tr>
<tr>
<td>5.3 Children passing tasks</td>
<td>35</td>
</tr>
<tr>
<td>6 Recent Developments in Theory of Mind Research</td>
<td>38</td>
</tr>
<tr>
<td>7 Additional Cognitive Accounts of Autism</td>
<td>39</td>
</tr>
<tr>
<td>7.1 Central Coherence</td>
<td>40</td>
</tr>
<tr>
<td>7.2 Executive Function</td>
<td>41</td>
</tr>
<tr>
<td>8 Language and Theory of Mind</td>
<td>43</td>
</tr>
<tr>
<td>8.1 A Shared Symbolic System</td>
<td>43</td>
</tr>
<tr>
<td>8.2 Impairments in Language associated with Autism</td>
<td>44</td>
</tr>
<tr>
<td>8.2.1 Semantics</td>
<td>45</td>
</tr>
<tr>
<td>8.2.2 Pragmatics</td>
<td>47</td>
</tr>
<tr>
<td>8.2.3 Narratives</td>
<td>49</td>
</tr>
</tbody>
</table>
2 Performance Across Tasks 94
   2.1 Participant Characteristics 94
   2.2 Measures 97
      2.1 Strange Stories 97
      2.2 Cartoons 98
      2.3 Eyes 98
   2.3 Controlling for chronological age, verbal age and IQ 98
   2.4 Overall view of advanced measures 100

Chapter 3: Discussion 103
1 Research Questions 103
   1.1 Are Advanced Measures Developmentally Appropriate? 103
      1.1.1 Standard Tasks 103
      1.1.2 Strange Stories 105
      1.1.3 Cartoons 107
      1.1.4 Eyes 109
      1.1.5 Advanced Measures 109
   1.2 Pass/fail or a Continuum of Mentalising ability? 110
      1.2.1 Mentalising abilities in autism 110
      1.2.2 Levels of mentalising complexity 114
      1.2.3 Individual Differences 116
   1.3 Are there characteristics associated with mentalising? 117
   1.4 Are cognitive and affective aspects of mentalising distinct? 119
      1.4.1 Types of mental state language 119
      1.4.2 Task loading of cognitive and affective skills 121

2 Methodological Issues and Clinical Implications 123
   2.1 Measures of Participant Characteristics 123
   2.2 Mentalising Measures 124
   2.3 Generalisability of findings 126
   2.4 Impairments in mentalising competence or performance? 129

3. Concluding Remarks 133

References 135
Tables

1. Participant Characteristics 84
2. Group Performance on Standard Theory of Mind 86
3. Group Performance on Advanced Theory of Mind Tasks 89
4. Pearson Correlations between Advanced Task Performance and Participant Characteristics 96
5. Pearson Correlations for Performance across Advanced Theory of Mind Tasks 97
6. Partial Correlations for Task Performance controlling for Chronological Age, Verbal Age and IQ 99
7. Discriminant Function Analysis of Variables 102

Figures

1. Histogram of Mean Complexity Scores for Individual Items on Strange Stories 91
2. Plot of Repeated Measures Analysis of Group Performance by Story Type 91
3. Histogram of Mean Scores for Individual Items on Cartoon Task 93
4. Plot of Repeated Measures Analysis of Cartoon Type by Group 93
5. Histogram of Mean Scores for Individual Items on Eyes Task 95
6. Z Scores for Group Differences across Advanced Measures 101
Appendices

1. Ethical approval 148
2. Information sheet for autistic group 149
3. Information sheet for typically developing group 150
4. Consent form 152
5. Scoring Criteria for the Strange Stories 153
6. Examples of Mentalising and Physical Stories 154
7. Scoring Criteria for the Cartoons 155
8. Examples of Mentalising and Physical Cartoons 156
9. Scoring Criteria for the Eyes 157
10. Examples of the Eyes 158
Abstract

The development of the ‘theory of mind hypothesis’ altered the autism research field by offering a single underlying cognitive impairment to account for the defining triad of impairments. Baron-Cohen, Leslie and Frith (1985) suggested that autism was the consequence of a deficit in the development of a theory of mind. This would mean that autistic children failed to develop an understanding that people have minds as well as bodies. Baron-Cohen et al. (1985) demonstrated that most autistic children, and typically developing children under the age of four years, were unable to state a story character’s mistaken belief, unlike typically developing children over 4 years who were able to do so. The finding was particularly significant because the theory of mind impairment appeared to be specific to autism, and not explained by intellectual or language ability.

False belief tasks developed to investigate autistic children’s ability to form mental representations of people’s thoughts, feelings and desires, used a pass/fail paradigm which was consistent with the view that theory of mind was an all or none affair. However, inconsistent performance by autistic children across a variety of false belief tasks, and the ability of some autistic children with good verbal ability to pass false belief tasks, challenged the model of an absence of any mentalising ability. Performance on false belief tasks at ceiling by many autistic children with a verbal age above 6 years has stimulated the development of more advanced tasks, scored on a performance scale to assess the extent of mentalising ability rather than its presence or absence.
This study matched 20 autistic children between 6-12 years with typically developing children for chronological age, verbal age and IQ. The children's theory of mind ability was assessed on 'standard' false belief and a picture sequencing task, and on three 'advanced' measures: an adapted version of the Strange Stories (Happe, 1994; Happe, unpublished), a children's version of the Eyes task (Baron-Cohen et al., 1997; Baron-Cohen, unpublished) and a new Cartoons task (Happe, unpublished). The autistic group's scores were significantly lower than typically developing children's in accounting for story characters' actions on mentalising stories, but not physical stories. The autistic group used significantly less mental state language in their answers than the typically developing group, with a specific difference in the use of second order phrases. Performance of the two groups was also distinguished by the Eyes task, but not the Cartoons task. Overall, the 'advanced' theory of mind tasks were able to discriminate group membership with 85% accuracy.

The use of alternative cognitive strategies to produce appropriate answers to lower level mentalising tasks in a more laborious way than the affective route used by typically developing children may explain these findings. In addition, it may be that these able autistic individuals experience an impairment in the ability to use low level mentalising abilities which they may in fact possess.
Acknowledgements

I would like to thank Francesca Happé and Simon Baron-Cohen for their collaboration and provision of materials for use in this study. Many thanks go to Tony Charman for his supervision and involvement at each stage of the research.

Special thanks go all the children and families who gave up their time to participate in this research, many travelling some distance to do so. I would like to thank Barbara Kugler, Sue Brown, Liz Relf and Clare Topping for their co-operation in setting up the study in their respective departments. My particular gratitude goes to Patti Rios and Fiona Cleaver for all their enthusiasm and help, particularly at difficult times.

In addition I would like to thank my family, particularly Rosalind and Michael Brent and Sasha and Marcus Sperber for their support, patience and practical help. Finally, many thanks go to Anna Picciotto, Liza Frank and Alex Datnow, without whom it would have been an even more stressful experience.
Introduction

1 Overview

The value of the theory of mind hypothesis was in drawing together a variety of threads from different fields of research to provide an integrated account of disordered development in linguistic, cognitive and social domains associated with autism (Baron-Cohen et al., 1993). Studies from the 1960s focused on language as a primary impairment (Hermelin & Conner, 1967), but preserved phonological and semantic skills (Tager-Flusberg, 1988) led to interest in the pragmatics of language (Tager-Flusberg, 1989).

It was hypothesised that children who were unable to develop a view of other people as mindful would remain unaware of the use of verbal and non-verbal communication to share thoughts and feelings (Tager-Flusberg, 1996). A failure to experience the innate pleasure of an interpersonal emotional connection with another mindful being might prevent the development of an awareness of other people's minds, thought to be necessary for the development of language (Hobson, 1993).

The theory of mind hypothesis suggests that children with autism do not develop an awareness that people have minds as well as bodies. This ability was thought to be acquired in typical development by the age of 4 years, demonstrated by the ability to pass a false belief task (Wimmer and Perner, 1983). Initially it was hypothesised that this ability was the result of a 'radical shift' in children's ability to form a representation of
other people's mental representations, described as meta-representations (Leslie, 1988). Studies of early interpersonal behaviour in typical development suggest that the ability to pass theory of mind tasks may actually be the culmination of a developing ability to think about desires, feelings and thoughts from the earliest months (Hobson, 1989).

The failure of the majority of autistic children to pass theory of mind tasks suggested that autism was the result of a deficit in the theory of mind mechanism (Baron-Cohen et al., 1985). Autistic children were described as ignorant of other peoples minds, which could account for their profound difficulty in social interaction and communication which depend on the ability to understand other people's intentions and interests (Baron-Cohen, 1989). However, the consistent finding that some autistic children pass what could be described as 'standard' theory of mind tasks (i.e. tasks currently used to assess low level first and second order theory of mind abilities) prompts the question whether autistic individuals do possess mentalising abilities, or implement an alternative problem solving approach (Happe, 1994a).

Tasks designed to assess children's theory of mind abilities can only be 'remote probes' for underlying cognitive processes, which may be confounded by linguistic and cognitive challenges (Frith, 1992). Studies have demonstrated links between both language and cognitive abilities, with task performance, but it remains unclear whether it is because these skills are necessary for the development of theory of mind skills or simply required for the non-mentalising aspects of the tasks. In an attempt to control for cognitive
demands of theory of mind tasks there has been a recent focus on autistic children with IQ within the normal range.

The good performance of more able autistic children has challenged the utility of the theory of mind hypothesis as an account of autism. If children can demonstrate mentalising abilities despite their obvious social and communication impairments an alternative explanation might be required. Proponents of the hypothesis suggest that good performance was the result of children with a verbal age over 6 years performing at ceiling on tasks designed for typically developing 3-5 year olds. This has stimulated the development of theory of mind tasks, which can be described as ‘advanced’ measures (Baron-Cohen et al., 1997), for these more verbally and intellectually able autistic children. These new tasks differ from standard measures in their use of a performance scale rather than a pass/fail paradigm.

This chapter will present an overview of the literature defining autism as a pervasive developmental disorder characterised by impairments in socialisation, communication and repetitive routines and interests (Wing and Gould, 1979). It will go on to present a critical review of research into the utility of the theory of mind hypothesis as an account of autism and outline additional approaches. Recent literature focusing on the association between theory of mind and verbal ability will be examined, tracing the possible origins of impairments beginning in infancy. Finally, this review will provide a rationale for the measures used in the study.
2 Autism

2.1 Historical Context

'Autism' is the diagnostic label used to describe a group of individuals presenting with pervasive impairments in communication, social interaction and a restricted range of repetitive behaviours and routines (Wing, 1988). The term 'autism' was introduced in the 1940s to describe children presenting with these unusual characteristics but it seems likely that this constellation of symptoms were observed before this time (Frith, 1989). The frequent co-occurrence of cognitive impairments with autism may explain why this particular pattern of symptoms were not identified earlier, lost under the umbrella of 'mental retardation.' Autistic individuals with language and intellectual abilities in the normal range may have had a diagnosis of schizophrenia (Happé, 1994b).

One particular case which has been suggested as being of a child who might now be diagnosed as autistic, lived in the woods of Aveyron (Itard, 1801). The particular factor that gave rise to this speculation was that he gained no language or social skills even after being taken into society and educated (Frith, 1989). This highlights the distinction between autism and an experience of severe social deprivation, which in itself does not prevent development of social and communication skills once in a nurturing environment (Clarke and Clarke, 1998). This supports the view that characteristics associated with autism are the result of an underlying organic impairment rather than an emotional response to environmental adversity (Rutter, 1970).
Bleuler (1908) is credited as the first to use the term autism, from the Greek ‘autos’ meaning ‘self,’ to describe the characteristics of a sub-group of schizophrenic patients who seemed withdrawn from the social world. ‘Autism’ was subsequently used to describe the socially isolated children seen independently by both Leo Kanner (1943) in the United States, and Hans Asperger (1944) in Germany.

Kanner (1943) was the first to write about children presenting with extreme social withdrawal, which he described as ‘autistic aloneness,’ who seemed to prefer relating to objects than to people. Asperger’s (1944) account published in the following year differed from Kanner’s in that he did not describe language impairments and commented on an avoidance of eye contact, peculiar facial expressions, stereotypic movements, narrow areas of interest and unusual speech patterns.

Characteristics of these children included distress resulting from a change in their usually rigid routines and sensitivity to alterations in their environment. Similarly their own behaviour seemed restricted to a limited range of activities which lacked creativity or imaginative play. Other characteristics included some good rote-memory skills, inconsistent with poor intellectual abilities, which tantalisingly suggested to Kanner some untapped potential. He also reported both heightened and dampened sensitivity to different stimuli, for example; apparently being startled by certain sounds while being non-responsive to being addressed.
Kanner (1943) emphasised that the impairments he had described in children continued into adulthood. Following this, however, there was a view that the disorder, initially referred to as 'early infantile autism,' might be cured or simply grown out of before adulthood. As children diagnosed with autism have become adults with ongoing specific needs, there has been a renewed emphasis on the importance of taking a developmental perspective of the disorder (Rutter and Schopler, 1988). The focus in this paper on children with autism should not preclude application to autistic individuals across the life span.

2.2 **Definition and Diagnosis**

Autism is currently described as a disorder of development, emphasising the importance of comparing features against the normal course of development. This also emphasises the distinction between the concepts of disorder and delay. Children who present with general delays in their progression through an expected pattern of social and cognitive development can be described as having a mental age that is inconsistent with their chronological age. However, the overall pattern of development would be consistent with a correspondingly aged typically developing child, whose chronological age matched their mental age.

This picture differs from that of an autistic child whose pattern of development does not appear to match a younger unimpaired child. For example, there are reports of autistic
children whose first words are semi-structured sentences and whose initial use of words is not followed by increased usage as is expected in typical development (Parke, 1987). Autism is characterised by qualitative differences in specific areas of development and is therefore described as a developmental disorder rather than general developmental delay.

Kanner (1943) proposed two defining features that he considered necessary and sufficient for diagnosis; autistic aloneness and an obsessive insistence on sameness. The former has continued to be the central criteria, encompassing a range of social and communication impairments. The concept of a preoccupation with sameness can be unpacked to describe a range of autistic characteristics, including repetitious behaviour and language, rigid routines and a narrow range of interests. The definition is based on observable characteristics, rather than on an aetiological description of the disorder, which at present remains unclear (Gagnon, Mottron and Joanette, 1997).

An epidemiological study was carried out to investigate the presence and coexistence of characteristics associated with autism in the London Borough of Camberwell (Wing and Gould, 1979). Following this study, three features were found to coexist as a syndrome rather than simply randomly co-occurring which have become the defining features of autism. Described as the ‘triad of impairments,’ these are: significant social difficulties, significant verbal and non-verbal communication difficulties, and a lack of imaginative skills replaced by a limited range of repetitive behaviours (Wing and Gould, 1979).
The Camberwell Study (Wing and Gould, 1979) investigated the population of learning disabled children in Camberwell in 1970. From the total number of 35,000 children under 15 years, 914 children were identified as known to health, social or education services. Within this group there were 173 children who were identified as having a learning disability and/or impairments in their social, communication or imaginative skills. If children presented with only one or two of these characteristics it would suggest that they were randomly co-occurring, while a consistent presentation of all or none would support the existence of a unitary syndrome.

Wing and Gould (1979) identified children whose intellectual disability was prohibiting the development of communication and imaginative skills as part of a global intellectual impairment. These skills normally develop by 20 months, so it would not be expected that children with a mental age below this would have developed them. All the children functioning above a mental age of 20 months who had social impairments were also characterised with communicative and imaginative impairments, whilst none of the children without social impairments presented with impairments in the other two domains. This supported the hypothesis that autism is a syndrome characterised by a triad of impairments (Wing, 1988).

The triad of impairments forms the basis of clinical diagnostic criteria (Rutter and Schopler, 1987), although specific classifications vary slightly. ICD-10 criteria include onset before 3 years old; qualitative abnormalities in reciprocal social interaction;
abnormalities in communication; and restrictive, repetitive and stereotyped patterns of behaviour, interests and activities, which are not attributable to other pervasive developmental disorders (International classification of diseases, 10th revision, ICD-10, WHO, 1990).

2.3 Epidemiology

2.3.1 Prevalence

Various studies have identified different levels of prevalence of autism according to how strictly diagnostic criteria are applied, between 5 per 10,000 (Fombonne, 1998) and 10 per 10,000 (Gillberg, 1998a). There is a suggestion that diagnosis is increasing which has led to concerns of over-inclusion, although the increase may result from increased awareness of the disorder associated with earlier and more reliable diagnosis (Bryson and Smith, 1998).

2.3.2 Gender Differences

A gender difference was evident in both Kanner (1943) and Asperger's (1944) early reports. The extent of this disparity varies according to inclusion criteria, between 2:1 (Ciadella and Mamelle, 1989) for studies using strict definitions of autism, and 5:1 for wider inclusion criteria (Lord and Schopler, 1987).

2.3.3 Intellectual Impairment

Various studies suggest that around three-quarters of autistic children have additional
learning disabilities, scoring under 70 on standardised IQ tests (Rutter, 1979). Interestingly this was not part of the picture presented by either Kanner’s or Asperger’s samples, probably because these children would then have been diagnosed as mentally retarded. The strength of the relationship between autism and learning disability has remained unexplained and is often treated as incidental, despite the predictive ability of IQ for long-term outcome (Bailey et al., 1996). Savant skills are frequently cited in the media when autistic persons demonstrate remarkable 'islets of ability,' such as accurately reproducing detailed images, sounds or numerical information (e.g. Sacks, 1995), although this savant profile is relatively rare.

2.3.4 Comorbidity

Epilepsy was recognised as elevated in autistic samples, affecting between a fifth to a third of individuals (Volkmar and Nelson, 1989). Unlike epilepsy associated with non-autistic learning disability, the onset of seizures is more common in late adolescence than in early childhood (Bailey et al., 1996). There is some evidence for a link between autism and Fragile X, Rett’s and Tourette’s syndromes (Barton and Volkmar, 1998).

2.3.5 Social Class

Kanner’s and Asperger’s samples were both characterised by children from families with an above average socio-economic status. Schopler et al.’s (1979) meta-analysis concluded that although differences in social class were identified in some of the studies, they were the result of other biases. Parents from middle class families were more likely to know
about autism and seek a diagnosis from specialists, resulting in an over representation in these groups. Since autism has been featured more prominently in the media there has been a decrease in this imbalance (Frith, 1989).

2.4 Course and Prognosis

Impairments associated with the triad of impairments are life long, although clinical presentation varies across the life span and between individuals (Sigman, 1998). Longitudinal studies suggest that approximately 10-15% of adults with autism have good outcome, 15-25% fair, 15-25% poor and 35-50% very poor outcome (Sigman and Capps, 1997). Compared to individuals with other language, social or developmental delays, individuals with autism, matched for age, IQ, verbal ability and demographic characteristics, have poorer adjustment in long-term follow-up studies (e.g. Cantwell et al., 1989). This research has also found that intellectual and language abilities are robust predictors for adult functioning. Those who make the greatest improvements have higher IQ (Lord and Schopler, 1989) and acquire language before 6 years (Szatmari et al., 1989). For those with relatively good verbal skills, the availability of appropriate support is an additional indicator of good outcome (Lord and Venter, 1992).

2.5 Differential Diagnoses

Diagnostic tools should facilitate differential diagnoses to be made by stating clear inclusion and exclusion criteria. Although diagnostic criteria reliably identify those falling within the range of pervasive developmental disorders from those who do not, they are
less clear for specific diagnoses within this relatively wide classification (Mahoney, et al., 1998). Clinicians refer to a ‘spectrum of disorders,’ that include individuals within a range of symptom severity, which will be referred to by the umbrella term ‘autism’ in this paper.

2.5.1 Schizophrenia

The diagnosis of autism partly arose from an increasing awareness of a group of children whose characteristics were not adequately described by their label of childhood schizophrenia (Rutter, 1978). Some accounts of schizophrenic negative symptomatology could overlap with some autistic behaviour, although positive symptoms show no similarities. While autism is usually diagnosed by 3 years, psychotic presentation rarely occurs before 6 years, after which shared characteristics seem more similar to schizophrenia than autism (Green et al., 1984).

2.5.2 Asperger syndrome

The most striking difference between Kanner’s and Asperger’s original accounts were the higher language abilities and normal IQ in Asperger’s sample. There is much debate, however, about whether those individuals presenting with fluent although unusual language are most usefully described as falling at the upper end of a continuum of autistic severity, or as a discrete diagnostic group (Frith, 1991). Wing (1981) was the first to use the term Asperger syndrome to describe individuals who developed language by three years of age but who demonstrated impairments on all the domains of the triad. The aim was to extend the classification to include individuals across a continuum of severity,
rather than to create a distinct diagnostic group.

Some consider Asperger syndrome to be distinct from high functioning autism on the basis of several distinguishing characteristics (Kugler, 1998). Asperger syndrome has been associated with near normal language by the age of six years, but with more verbose speech (Klin, 1994). Cognitively, children with Asperger syndrome fall within the normal range, with psychometric profiles characterised by higher performance on verbally mediated sub-tests and lower performance on performance based sub-tests, compared to the reverse pattern associated with high functioning autism (Klin et al., 1995).

Prior and her colleagues' (1998) cluster analysis identified three sub-groups, loosely matching diagnoses of autism, Asperger syndrome, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). However, they were found to share characteristics, distinguished only on the basis of severity. This was consistent with the paradigm of a continuum of ability rather than the existence of distinct clinical groups (Prior et al., 1998).

The ICD-10 and DSM-IV classification systems both use a hierarchical rule that characteristics meeting the criteria for autism preclude a diagnosis of Asperger syndrome. Most individuals on the continuum with normal IQ and minimal delay in language fit the criteria for autism making the Asperger classification redundant (Gillberg, 1998b).
2.5.3 *Semantic Pragmatic Disorder*

Semantic Pragmatic Disorder (SPD) was described by Rapin and Allen (1983) as one of six developmental language disorders, characterised by expressive and receptive impairments in the use of language to communicate in a meaningful and appropriate way. An additional diagnostic group would only be of use if there was a discrete group of children with pervasive communication difficulties, without impaired social and imaginative skills (Boucher, 1998). Children diagnosed with SPD do however appear to have poor imaginative play (Bishop, 1989) and impaired social and non-verbal communication difficulties (Shields et al., 1996a). SPD would therefore seem to be a duplication of Asperger syndrome and/or high functioning autism (Shields et al., 1996b).

2.6 *Causal Theories*

Autism is currently diagnosed and defined according to clinical characteristics without an aetiological explanation. The disorder was previously thought to be a form of childhood psychopathology with a psychogenic origin until Rutter (1970) identified it as a form of organic brain damage. Bettelheim (1956) had suggested that autism was a defensive response to the absence of a secure attachment to a cold mother and recommended the removal of the child to a residential setting.

The biological evidence for an organic impairment in autism comes from several levels of explanation (Bailey et al., 1996). An early twin study by Folstein and Rutter (1977) identified a genetic association with the disorder which has since been estimated at
between 3-7% for siblings. This is a 50-100 times greater risk than in the total population (Jorde et al., 1991). There is evidence to support the inheritance of a broader phenotype for autism, characterised by a less severe range of social and cognitive impairments associated with normal intelligence, but which is not associated with epilepsy (Bolton et al., 1994). The mode of genetic transmission might be as few as two to five genes, indicated by the significant drop in concordance rates from monozygotic to dizygotic twins (Bailey et al., 1995), and from first to second degree relatives (Jorde et al., 1991).

Possible abnormalities have been examined at a neurobiological level, including some evidence of abnormal levels of serotonin in the brain (McBride et al., 1989). There have also been abnormalities found in the limbic system and cerebellum, but these may be the result rather than cause of other brain abnormalities associated with autism (Kemper and Bauman, 1998). There seems to be more support for diverse neuroanatomical irregularities, associated with increased brain size, rather than for a localised area of abnormality in the brain (Bailey et al., 1996).

Bailey and his colleagues (1996) argued for an integrated approach to these potentially complementary models that have so far developed independently of one another. They emphasised that there may not be a single underlying aetiology but rather a complex interacting system acting at the biological, cognitive and behavioural levels. A model suggested by Morton and Frith (1995) was the combination of various biological abnormalities resulting in a single cognitive deficit, which in turn produced a variety of
behavioural sequelae. As a consequence of individual differences at the biological and behavioural levels, they suggest the most useful focus for research would be at the cognitive level.
3 The Theory of Mind Hypothesis

Irrespective of the intellectual impairments experienced by three-quarters of autistic individuals, the disorder is characterised by specific cognitive abnormalities across the ability range. It was hypothesised that they might be indicative of an underlying cognitive dysfunction that could in turn account for the clinical features associated with autism.

Initially the cognitive impairments were associated with difficulties in abstracting meaning from various linguistic stimuli (Hermelin and O’Conner, 1970). Although these difficulties might reveal impairments in language development, they could not fully account for the social impairments central to the disorder (Bailey et al., 1996). The focus shifted to the arena of social cognition, at a time when Baron-Cohen’s experience of teaching children with autism left him feeling that “they appeared singularly oblivious to what others were thinking” (Baron-Cohen, 1992, p. 9, author’s italics).

We use our understanding of people’s thoughts, beliefs, attitudes, motivations and intentions to understand and predict their behaviour, which is in turn necessary for social interaction and communication. If people with autism had a specific cognitive impairment which prevented them from understanding the contents of people’s minds, a ‘mindblindness,’ it might account for the triad of impairments associated with the disorder (Baron-Cohen, 1995). Without the ability to interpret the otherwise apparent chaos of sounds and movement in their environment, it is perhaps not surprising that
autistic individuals cling to rigid and repetitive, but reassuringly predictable, routines (Baron-Cohen, 1992).

Any useful account of autism must explain impairments compared to the pattern of typical development and aid differential diagnoses. Such an account must also offer testable predictions regarding areas of specific impairments and of preserved skills (Frith, 1992). Such a claim has been made about the ‘theory of mind’ hypothesis which was initially a model for the acquisition of social cognition in typical development. The term described the ability to understand that others have different thoughts from one’s own (Premack and Woodruff, 1978). It was described as a theory, “first because such states are not directly observable, and second because the system can be used to make predictions, specifically about the behaviour of other organisms” (Premack and Woodruff, 1978, p.515).

3.1 False Beliefs

In order to explore the development of a theory of mind there has been extensive research into children’s ability to predict other people’s beliefs, and how these beliefs might influence their behaviour. In order to assess children’s understanding of other people’s thoughts it would be most useful to ask them about thoughts that differ from their own, such as false beliefs. This enables a distinction to be made between children’s ability to describe their own knowledge and their understanding the thoughts of others (Baron-Cohen, 1992).
Wimmer and Perner (1983) investigated at what stage typically developing children acquired this ability. They developed a series of stories involving characters who had left objects in certain places and did not see that they had subsequently been moved. Children were shown a story with dolls about Maxi, who was told by his mother to put his chocolate in a blue cupboard. While he was out she moved it to the green cupboard and then left the house before Maxi returned. Participants were asked to say where they thought Maxi would look for his chocolate when he came home. In order to pass the test, children would have to think about Maxi's limited information and understand that he would think the chocolate was where he had left it. The correct response would be to say that Maxi would look in the blue cupboard, even though the participants themselves knew that it was really in the green cupboard (Wimmer and Perner, 1983).

Wimmer and Perner (1983) found that only 15% of their typically developing 3-4 year old participants passed, while almost 80% of 4-6 year olds were able to attribute false beliefs to the story characters, and 6-9 year olds passed easily. They concluded that there seemed to be an "abrupt developmental step at the age of four" (Wimmer and Perner, 1983 p. 125). Wimmer and Perner (1983) did not think this change was simply the result of increasing cognitive capacity and proposed a radical conceptual change in the way children represent the world around them at this age.

Young children develop views about the world based on exact replications of reality in
their mind, rather than having an awareness that people’s knowledge is based on a mental construction of what is seen and heard (Wellman, 1990). The acquisition of a theory of mind around four years old may be the result of the developing ability to distance reality from a mental representation of it, and therefore acquire an openness to the possibility of alternative interpretations of the environment (Baron-Cohen, 1995). Leslie (1988) described this as dependent on the switching-on of an innate ‘decoupling mechanism’ enabling the formation of ‘metarepresentations.’ A decoupled representation is composed, for example, of an agent; Johnny, an informational relation; pretends, an anchor; the banana, and a decoupled expression; is a telephone (Leslie, 1988).

3.2 “Radical shift” or “Matter of degree” in the acquisition of theory of mind?

Leslie’s (1988) view of a radical shift, facilitated by the ability to generate a metarepresentation, has been challenged by those who propose a gradual development in children’s understanding that other people’s construction of reality is based on their interpretation of the world around them. Wellman (1990) suggested that from around three years old, children begin to be aware of the difference between real and imagined things, such as a real biscuit being visible and edible whilst an imagined one is not (Wellman and Estes, 1986). However, it is not until around the age of four that children realise that others would believe that a sponge painted to look like a rock was a rock, even though they themselves had been shown it was a sponge (Flavell et al., 1983; Gopnik and Astington, 1988).
The recognition that the development of a child's theory of mind may not simply be all or nothing has opened the area up to the subtleties of children's performance on experimental tasks and children's underlying skills. Frith (1992) acknowledged that false belief tasks could only be 'remote probes' to underlying cognitive abilities, but what has become increasingly apparent is the variety of factors which come into play when a child performs even a simple task. Initially failure on a false belief task was taken to indicate the absence of theory of mind (Leslie, 1988) but it now appears it may only serve to indicate that a certain threshold of ability has not been passed.

An increasing number of studies have demonstrated that slight alterations in theory of mind task presentation (e.g. Saltmarsh et al., 1995) increase the proportion who can pass at three years. This does not suggest that children do possess a developed theory of mind below four years which is masked by the task format, but it does highlight the importance of exploring factors which impact on the assessment of children's abilities. This will have particular relevance for our understanding of the meaning of theory of mind task performance for children with autism.
Baron-Cohen, Leslie and Frith (1985) discovered that although around 85% of non-autistic (typically developing and Down's syndrome) children with a mental age above 4 years passed a false belief task similar to Wimmer and Perner's (1983), 80% of autistic children could not. The autistic children in this study ranged between 6 and 16 years, with a mean verbal age of 5 years 6 months; the children with Down's syndrome were of a similar chronological age but with poorer verbal ability; while typically developing children had a mean age of 4 years 6 months.

The modified task that was used was called the Sally-Ann task, in which two dolls, one with a box and the other with a basket, were presented to participants. Sally was first shown to put her marble in her basket and then left the scene. While she was gone, Ann moved the marble from the basket into her own box. Once Sally returned, children were asked to predict where Sally thought her marble was. Participants needed to demonstrate an understanding that although they knew the marble was in the box, Sally had not seen the transfer and would therefore think the marble was in the basket where she had left it (Baron-Cohen et al., 1985). The mistake made by the autistic children was to point to the real location of the marble when asked where Sally would think it was.

The task incorporated control questions to minimise the possibility that impairments were the result of non-mentalising difficulties. Autism is associated with specific
communication impairments so it would not be surprising if autistic children’s difficulty on this task was a result of problems in understanding the experimenter’s question. In order to test if children understood the language of the test they were asked, “Where is the marble really?” A memory question checked whether errors were because children simply could not remember the location of the marble, “Where was the marble at the beginning?” The children with autism passed these questions, which was taken to demonstrate that their impairments were specifically related to the mentalising question (Baron-Cohen et al., 1985).

Another strength of the argument was the autism-specific nature of the impairment in predicting Sally’s thoughts, rather than a more general cognitive impairment which would again be unsurprising considering that the majority of children with autism present with additional learning disabilities (Baron-Cohen et al, 1985). Subsequent studies have supported the finding that other clinical groups have little difficulty with these tasks, such as non-autistic children with learning disabilities (Baron-Cohen, 1989), language difficulties (Perner et al., 1989), conduct disorder (Happe and Frith, 1996) and ADHD (Charman et al., 1999).

The acquisition of theory of mind was initially considered an all or none ability, which typically developing children were expected to acquire at approximately four years old (Leslie, 1988). Eisenmajer and Prior (1990) presented children with two false belief tasks and found that only 3 of the 29 autistic participants did not perform consistently across
both. Happé (1995), however, found a proportion of autistic participants who did perform differently across parallel forms. This was evidence to support a growing interest in factors that might mediate children’s ability to perform on theory of mind tasks.
5 Criticisms of the Theory of Mind Account

5.1 Imaginative impairments

There have been criticisms of the claim that children's performance on theory of mind tasks provides evidence that the triad of impairments can be accounted for by a metarepresentational deficit. De Gelder's (1987) criticism of the false belief tasks was that children with a specific impairment engaging in play might have difficulty in a task requiring children to 'imagine' motives and beliefs in a doll that actually has no mind at all.

Leslie and Frith (1988) presented children with an alternative false belief task that was not dependent on imaginative play. The experimenters asked children what they thought was in a Smarties tube they were shown. Children usually responded by saying "Smarties!" The experimenters then opened the tube and showed participants they had in fact put a pencil inside instead. The lid was replaced and children were asked what another person coming into the room, who had not seen inside the tube, would think. The majority of typically developing children over the age of four years predicted that someone else would say "Smarties," although children with autism, and under 3 year old typically developing children said what really was inside, "a pencil."

Saltmarsh et al. (1995) suggested that younger children found it easier to think about concepts which have a physical quality than abstract ones. In the Smarties tube task,
children under 4 years found it easier to think about the pencil they had seen in the tube, rather than the Smarties they simply imagined had been in the tube. Showing children the Smarties being taken out of the tube at the beginning of the task (Wimmer and Hartl, 1991), or posting a picture of what they thought was in the tube (Mitchell and Lacohee, 1991), provided a physical reality which enabled younger children to acknowledge their prior false belief (Mitchell, 1997).

This further supports the view that the type of task and manner in which it is presented to autistic or typically developing children influences their ability to access some level of mentalising ability, even if it may not be as developed as in non-autistic or older children. It is important to emphasise that modifications to tasks which increase the proportion of children able to pass, do not reveal an intact ability to understand others’ minds, but do highlight the presence of some ability to think about others’ minds in young and autistic children. This contrasts with the view that a theory of mind is either present or absent, or suddenly switches on at a certain stage. It also raises the question of what other factors facilitate or inhibit children’s performance on these tasks.

5.2 Language impairments

Children with autism are characterised by a pervasive impairment in verbal and non-verbal communication. This suggests they might not understand the language used in the theory of mind tasks. Baron-Cohen et al. (1986) developed a comic-strip story-sequencing task to identify whether children could understand a simple story without
words. Children were then asked to tell the story of the pictures, to provide more information about any differences in narratives provided by autistic compared to non-autistic participants (Frith, 1989).

The study used three types of story in order to explore whether difficulties ordering sequences were specific to those involving an understanding of mental states. Mechanical stories showed a sequence of physical causal relationships, such as a boulder having been pushed down a hill. Behavioural stories were based on social activities that did not rely on an understanding of story characters' thoughts, such as a girl stealing a boy's ice cream. Sequences described as mentalising were those in which understanding a character's implicit thoughts was necessary for a full understanding of the story, such as a boy not realising there was a hole in his bag out of which his sweets were falling.

Autistic children performed similarly to controls on behavioural and mechanical items, but significantly worse on the sequencing of mentalising stories. In the narratives, autistic children used significantly less mentalising language than typically developing or Down's syndrome controls (Baron-Cohen et al., 1986). Rather than a general difficulty, children with autism appeared to have a specific problem with the mentalising items, both for non-verbal processing of the story sequence and in producing mentalising language.

5.3 Autistic children who pass tasks

A weakness of the theory of mind hypothesis is the absence of a convincing account of
the anomaly that 38% to 60% of autistic children pass false belief tasks (Dahlgren and Trillingsgaard, 1996). One explanation could be that children who pass were misdiagnosed and are not really autistic, but this would simply establish a circular argument that children failing theory of mind tasks were autistic because they failed theory of mind tasks (Mitchell, 1997). This is also an unsatisfactory argument as those who pass the tests are still clearly impaired in real life interactions (Bowler, 1992).

In response, Baron-Cohen (1989) suggested that although some autistic children were able to pass very simple false belief tasks they did not tap the cognitive skills required for the complex understanding of real life experiences. He described the skills assessed in Sally Ann and Smarties tasks as ‘first order false belief tasks,’ because they were based on an understanding of what person A thinks. More complex situations were developed which assumed an understanding of what person B thinks person A thinks, described as second order theory of mind. Typically developing children aged 6-7 years are able to predict someone’s belief about a third person’s knowledge (Perner and Wimmer, 1985).

Baron-Cohen (1989) hypothesised that those autistic children who were able to pass first order false belief tasks would fail on second order tasks. Children were presented with a rather complex story about an ice cream van, beginning with two characters being told that the ice cream van would be in the park all day. Later on each character is told individually that in fact the van will be at the church, not knowing the other has received this information. Participants have to predict that one story character would think that the
other one was ignorant of the change (i.e. holds a false belief about another person's belief).

Initially it appeared that children with autism who were able to pass first order theory of mind tasks were not able to pass this more advanced test, unlike typically developing children (Baron-Cohen, 1989). However, subsequent studies have demonstrated some autistic individuals can also pass these second order false belief tasks, particularly by participants with verbal intelligence within the normal range (Bowler, 1992; Ozonoff, et al., 1991). Research suggests that verbal mental age was a discriminating factor between passing and failing theory of mind tasks by autistic participants (Dahlgren and Trillingsgaard, 1996).
Recent Developments in Theory of Mind Research

An explanation was needed to account for the autistic individuals who passed first and second order theory of mind tasks but who still presented with pervasive social impairments in the real world. In order to represent more realistic situations in which autistic children had difficulties, Happé (1994a) developed a series of scenarios in which characters said things which they did not mean literally, such as jokes, double bluffs and white lies, and presented them to autistic and non-autistic children.

Happé (1994a) found that the autistic children's ability to understand characters' behaviour in the Strange Stories correlated with their performance on false belief tasks. She concluded that the task seemed to tap the same underlying construct as false belief tasks. However, this more complex task did demonstrate that even able autistic participants differed significantly from matched controls by giving fewer explanations which referred to characters' mental states (Happé, 1994a).

Baron-Cohen et al. (1997) argued that passing false belief tasks should not necessarily imply the presence of intact theory of mind skills. False belief tasks were originally developed to assess developing abilities in three to six year olds, and were therefore likely to be too simplistic for older autistic individuals who have verbal IQ scores within the normal range. Baron-Cohen et al. (1997) devised a more advanced test of theory of mind ability, asking participants to choose a mental state word to describe someone's eyes.
A strength of the theory of mind hypothesis has been to account for the triad of impairments, which are central to diagnostic criteria (Wing and Gould 1979; Rutter and Schopler, 1987). However, a criticism has been its limitations in addressing secondary features commonly associated with the disorder. Various studies found abilities which distinguish autistic from non-autistic children, such as the former completing jigsaws by shape rather than by picture (Hermelin and Frith, 1991) and sorting faces by accessories rather than by expression (Weeks and Hobson, 1987), which remain unaccounted for by the theory of mind hypothesis.

In psychometric testing children with autism typically show an erratic performance profile across WISC sub-tests, in contrast to the more consistent performance of typically developing children (Frith and Happé, 1994). Autistic children perform poorly on those sub-tests involving communication skills and better on those requiring an ability to detach from contextual meaning and focus on detail, such as Block Design (Shah and Frith, 1993). This pattern highlights the characteristic of striking islets of ability; and an obsessive desire for sameness, which were included in Kanner's and Asperger's descriptions accounts.
7.1 Central Coherence

Frith (1989) suggested that autistic children had a specific difficulty in drawing together information from different sources to process meaning from context, which she described as central coherence. In contrast, typically developing children seem predisposed to make sense of stimuli at the expense of detail, such as remembering the meaning of stories rather than the words used (Hermelin and Frith, 1991). This accounts for autistic children’s superior performance on Block Design and Embedded Figures Tests (Shah and Frith, 1983) and resistance to optical illusions (Happe, 1996), which are explained by their tendency to disengage from the larger pattern/picture in favour of detail.

Happe (1997) explored the relationship between children’s drive for coherence and ability on theory of mind tasks. She found that autistic children were alike in their difficulty in using contextual meaning to help them choose the correct pronunciation of homographs in sentences, irrespective of their varying ability to pass first and/or second order theory of mind tasks. This is important in accounting for the enduring social difficulties experienced by even able autistic individuals in the real world despite their ability to pass second order theory of mind tasks (Bowler, 1992). It may be that without central coherence, autistic individuals are unable to use the even the limited mentalising abilities which some may in fact possess (Frith and Happe, 1994).
7.2 Executive Function

Repetitive behaviours and narrow areas of interest from which individuals find it difficult to detach themselves are characteristics associated with autism which are also unaccounted for by the theory of mind hypothesis. Interestingly, patients recovering from frontal brain damage have been observed to have similar difficulties associated with switching attention between different demands. In addition, both autistic individuals and patients with frontal lobe damage frequently use formal language and often have difficulties with social cues regarding reciprocity (Joanette et al., 1990). Abilities associated with the frontal region of the brain are those responsible for higher order skills such as shifting attention and organising abstract plans, which are described collectively as executive functions.

Following the observation of these similarities between autism and frontal lobe patients, tasks designed to assess executive dysfunction have been presented to both. The Tower of Hanoi requires participants to plan how to reassemble rings of decreasing diameter from one peg to a third peg, through several intermediate stages. The Wisconsin Card Sort allows participants to arrange cards according to different categories (colour, shape and number) with the intention that they should shift their attention to another categorising system once told they have made a mistake. Unlike typically developing controls, people with autism had difficulty with these tests irrespective of performance on false belief tasks, and despite the absence of any mentalising component (Ozonoff et al., 1991).
Russell and his colleagues (1991) suggested that the impairments associated with autism were the result of difficulties inhibiting impulsive responses associated with executive dysfunction. They demonstrated that like three-year-olds, children with autism were unable to learn that if they pointed to an empty box instead of one containing chocolate, they would actually get the chocolate for themselves (Hughes and Russell, 1993). They concluded that children failing false belief tasks did not have a specific difficulty mentalising, but a difficulty inhibiting the expression of their own true belief (Russell et al., 1991).

Mitchell (1997), however, argued that although executive function and mentalising difficulties may coexist they were not part of a single underlying deficit. He drew on evidence that children with autism pass a false photograph task in which an instant picture is taken in front of the child of a doll on a bed. Before the photograph has time to develop, the experimenter moves the doll onto the mat and then asks the child where the doll in the photograph will be (Zaitchik, 1990). Following Russell et al.’s (1991) argument, children would be expected to impulsively say where the doll currently was, but in fact autistic children were able to answer correctly (Leekham and Perner, 1991; Leslie and Thaiss, 1992), although three year old typically developing children were not (Zaitchik, 1990). Mitchell (1997) concluded that executive function could not provide a full explanation for mentalising deficits, and proposed children with autism might be characterised by both mentalising and executive function difficulties.
8 Language and Theory of Mind

8.1 A Shared Symbolic System

Once children become aware that people have minds which differ from their own, they seem motivated to find a way to share their own thoughts and find out what other people think (Hobson, 1984). This step necessitates the development of language. Language is a symbolic system that depends on the ability to detach a real object from an arbitrary but agreed word representing it. For example, the word "hat" represents a group of objects with shared characteristics, such as being worn on the head, and includes specific types such as "a cap" and "a top hat," but excludes other possibly related objects such as a "scarf" or a "bucket" (Tager-Flusberg, 1985).

A language system evolves from a consensus, enabling the speaker to express him/herself and the listener to understand. Hobson (1993) suggested that an infant's sudden increase in word learning is associated with the realisation that not only can words be used to describe objects but that everything already has an agreed name of its own. The motivation to express oneself and use language seems reinforced by the realisation that communication has an impact on the listener, acting to engage people's minds in an innately pleasurable way (Hobson, 1989).

Language development in turn may facilitate children's ability to self-reflect and become self-conscious (Hobson, 1993). For example, a child has to learn that others describe the
feeling that results in a smile as being 'happy.' S/he does not think about being happy and then learn to smile in order to let other people know. This suggests an interaction between language development and a growing awareness of mental states (one's own and other people's) during infancy, rather than a simple cause-effect relationship.

8.2 Impairments in language associated with autism

Early definitions viewed autism primarily as a language disorder because of the significant difficulties in speech production and comprehension (Rutter, 1978). Many children at the lower functioning end of the spectrum present as mute, but for those who develop language it is usually delayed, characterised by unusual qualities, and not compensated for by alternative communicative gestures (Loveland and Landry, 1996).

Similarly to other areas of impairment, there is a disordered pattern of development rather than a general delay across all domains. Autistic children present with some preserved linguistic abilities, such as lexical, grammatical and phonological skills (Tager-Flusberg, 1981). Although these skills are usually still delayed in comparison with typical development, they are similar to non-autistic children with matched verbal age (Tager-Flusberg, 1989). It seems that the areas which are less impaired are those which do not require the shifting of perspective dependent on speaker/listener attitudes, that is, those which do not depend on an intact theory of mind (Happe, 1994c).
In an analysis of spontaneous speech over a year, Tager-Flusberg and Anderson (1991) found that at the earliest stages of language development, autistic, Down's syndrome, and typically developing children matched for verbal ability, had a similar pattern of language abilities. This suggested an initial pattern of delay rather than disorder for autistic children just beginning to develop language. Over the course of a year, however, there was an increasing disparity between autistic and non-autistic groups in their ability to respond to others in a topic related way, despite demonstrating developments in structural aspects of their language (Tager-Flusberg and Anderson, 1991). Again, it was apparent that the specific areas of disordered language development were those which depended on an understanding of mental states. Unlike autistic children, typically developing children develop an understanding that language is used to share relevant ideas with an interested listener (Tager-Flusberg, 1996).

8.3.1 Semantics

An innate motivation to develop language in order to convey ideas between people is consistent with an idea that minds are primed to meaning. This has been demonstrated by typically developing children's ability to remember more words if they are meaningfully related, than a string of random words (Hermelin and O'Conner, 1967). Conversely, children with autism do not seem to show any difference in their ability to remember semantically related or random word strings and do not apparently draw on the meaning of words to help recall (Tager-Flusberg, 1991).
One explanation for this might be that children with autism are impaired in their ability to process word meanings and mentally organise them in a useful way. Tager-Flusberg (1985) explored this possibility by asking children to identify specific objects as either within or outside particular conceptual classes. For example, if the category was "boat," participants were asked to classify "canoe," "raft," "buoy" and "kite" as part of, or excluded from, the category. However, there were no significant differences between autistic and non-autistic participants' ability suggesting this was not the source of the impairment (Tager-Flusberg, 1985).

When autistic children are given a semantic cue to aid recall, their performance on memory tasks is comparable to matched controls (Tager-Flusberg, 1991). Since they are able to process word meanings, it seems that the source of the impairment is their ability to use available semantic information in the free recall task. Consistent with this hypothesis was the finding that children with autism were significantly less likely than matched controls to use the meaning of a sentence to choose the most appropriate pronunciation of an ambiguous word in a homograph reading test (Happe, 1997). It seems that for some skills, even where autistic children do have some underlying ability, it may not be demonstrated by task performance that requires participants to draw on knowledge from previous experiences, which typically developing children do quite spontaneously.
8.3.2 Pragmatics

There is mounting evidence suggesting a significant deficit in the appropriate use of skills according to the context in which autistic people find themselves. Pragmatic language skills are those which relate to the use of language as a tool with which to communicate with other people. Communication implies not only the speaker’s ability to express ideas but also the intention to convey a message that will be understood by, and relevant to, the listener. If pragmatic skills are responsible for the application of linguistic rules to a social context, the foregoing evidence might suggest these would be skills most impaired in autism. There are several patterns of pragmatic irregularities that have been identified in descriptions of autistic individuals, including echolalia, pronoun reversal, formal and literal speech patterns and a lack of prosody (relating to the rhythm and intonation of speech).

Echolalia describes the tendency of autistic individuals to repeat other people’s questions or phrases without modifying them to the current context. For example, a child requesting a drink might ask, “Do you want a drink?” rather than “Can I have a drink?”, perhaps because this was the original phrase associated with receiving a drink. There seems to be a difficulty using language flexibly, applying units of language in a new way to make sense in the new speaker/listener roles (Baltaxe, 1977).

In typical development, personal pronouns usually appear around the end of the second and the beginning of the third year, at the same time that children start using mental state
language (Hobson, 1993). Children with autism appear to have a particular difficulty with the use of personal pronouns, mixing up I and you, and often using proper names even to refer to themselves (Jordan, 1989). Experimental studies suggest that autistic children do not differ from typically developing controls using and understanding personal pronouns in structured settings, supporting the view of impairments in their unstructured use (Lee et al., 1994).

Prosodic characteristics, such as unusual intonation, pitch and rhythm, suggest autistic children are not aware of how they sound to other people, and how to make their speech sound interesting. Even fluent autistic speakers have difficulties initiating or ending conversations, or judging an appropriately long response. This is further evidence to support the view that autistic language impairments are in the subtle modification of general linguistic rules appropriate to different situations (Baltaxe, 1977).

Autistic individuals’ difficulty in using language to communicate can be reflected by the pedantic use of formal language. There are particular difficulties with metaphorical language which requires the understanding of the speaker’s intent (Happé, 1993). Conversation can be limited to narrow areas of the autistic person’s interest with no recognition of what might be boring or even intrusive to the listener, or what information a listener does/does not need in order to make sense of the conversation (Loveland et al., 1989).
8.3.3 Narratives

Narrating stories, at which typically developing children are skilled at around the age of six years, requires different abilities to those used in spontaneous conversation (Peterson & McCabe, 1983). Children with autism have difficulties in theory of mind which impair their ability to understand listeners' needs and interests, and this has implications for their ability to narrate stories (Loveland et al., 1990). Tager-Flusberg (1995) found autistic children's ability to tell coherent stories correlated with their performance on theory of mind tasks. However, theory of mind performance did not correlate with narrative complexity or the use of mental state words for the non-autistic participants (Charman and Shmueli-Goetz, 1998). This might suggest that autistic children's development of theory of mind and pragmatic language skills are more interdependent than for typically developing children.

8.3 Mental State Language

The subjective nature of understanding other people's minds makes it difficult to assess the beginnings of a child's awareness of mental states. The false belief tasks were used as an attempt at identifying the onset of these skills, but evidence suggests that children will have already developed an understanding of the concept of minds by the time they can pass such tests. In order to identify the beginnings of this skill, there has been interest in the time children first use mental state language in their spontaneous speech (Tager-Flusberg, 1992). Children with abnormalities in their ability to understand minds would be expected to show impaired emergence of these mental state terms (Hobson, 1990b).
Naturalistic studies using recordings of children’s spontaneous speech at home with their families, suggest that even before two years of age they are aware of the concept of desire, expressed by the phrase, “I want it” (Wellman, 1990). Emotion words referring to internal mental states, such as “sad,” “tired” and “hungry,” appear during their second year to refer to both children’s own and other people’s feelings (Bretherton et al., 1981). The development of more complex mental state language, such as “remember,” “think” and “know,” emerges during the third year (Bartsch and Wellman, 1989). However, some mentalising statements may be used as colloquial expressions without representing a real understanding of people’s minds, such as when “you know?” is used as a pause-filler (Shatz et al., 1983).

Autistic children can link simple emotion words with behaviour, and can predict story characters’ behaviour based on knowledge of their desire (Tan and Harris, 1991). In contrast they use significantly fewer words relating to attention and cognition than matched children with Down’s syndrome (Tager-Flusberg, 1992). This suggests a distinction between autistic children’s use of the earlier emerging words associated with desire and emotion, and those relating to cognition, which appear later in typical development (Hobson, 1990b).

Rather than focusing on the ability to label simple emotions, Hobson et al. (1989) argued that an impairment in the understanding of the way minds can engage with one another
may be more apparent in complex non-emotion words. For example, words such as 'friend' require an experiential understanding of what it is like to be or have a friend, while it may actually be easier to learn that when people shed tears, or cannot find something they want, they are described as 'sad.'
There seems to be a robust correlation between 'standard' false belief task performance and verbal mental age (Happe, 1995; Dahlgren and Trillingsgaard, 1996) but not with non-verbal measures of intelligence (Happe, 1993). It might be suggested that children who fail tasks are not able to understand the concepts used, rather than have specific deficits in mentalising. Alternatively, linguistic ability could be a secondary consequence of a less impaired theory of mind.

Happe (1995) found that the lowest ability at which a typically developing child could pass first order false belief tasks was 2 years 10 months, with a chronological age of 3 years 8 months, and a verbal IQ of 88. All normally developing participants above 6 years 9 months passed theory of mind tasks. In the non-autistic learning disabled group, the lowest range of ability was a verbal mental age of 3 years 10 months, chronological age of 11 years 11 months, and verbal IQ of 40. Interestingly, in the learning disabled group there was no relationship between verbal mental age and task performance.

In comparison, the lowest ability passer from the autistic group had a verbal mental age of 5 years 6 months, chronological age of 14 years, and a verbal IQ of 55. It seems that in order to pass theory of mind tasks, children with autism require significantly more developed verbal skills than non-autistic children. All participants with a verbal mental
age below 5 years 6 months failed, while all with verbal mental age above 11 years 7 months passed (Happe, 1995).

The upper threshold for verbal mental age in the autistic group could be explained in at least two ways. Either that theory of mind tasks can be performed by anyone with verbal competency above a certain level, or that without theory of mind individuals will not achieve a verbal mental age above 11 years 7 months (Happe, 1995). It appears that a certain level of verbal ability is a necessary but not sufficient skill for the development of theory of mind (Charman and Baron-Cohen, 1992).

An upper threshold reveals that some autistic individuals can develop the ability to pass false belief tasks, which could either indicate a delayed acquisition of a theory of mind or suggest that some had developed an alternative strategy to “hack out” solutions in more of a problem solving way (Frith et al., 1991). The latter explanation would be supported if, despite good task performance, individuals still experienced profound social impairments in more complex and subtle real life settings. In contrast, if individuals who passed false belief tasks had really developed an intact theory of mind, they would be expected to demonstrate better social skills in real life settings (Frith et al., 1994). A combined study of false belief task performance and social functioning identified two subgroups of individuals passing false belief tasks, one with raised social abilities and one without, consistent with each of the proposed explanations (Frith et al., 1994).
In an attempt to control for language difficulties, non-mentalising questions were included in false belief task studies (Baron-Cohen et al., 1985). However, physical control items have been criticised for being more concrete and therefore simpler than the experimental questions, allowing participants to perform at ceiling. Hobson and Lee (1989) investigated the relative complexity of words in the BPVS (Dunn et al., 1982), which researchers blind to the study's hypotheses rated for emotion and abstractness. Comparing the profiles of matched samples, they found that children with autism had a specific difficulty with mentalising words but actually performed similarly to matched controls for non-mentalising abstract words rated as equivalent in complexity. This casts doubt on the explanation that autistic children's performance can be explained simply by linguistic complexity in task presentation.
10 Early characteristics associated with autism

There is increasing interest in the child's gradual development of mental state awareness during infancy and its implications for the onset of impairments associated with autism. Rather than suddenly occurring at 4 years of age, the ability to pass false belief tasks seems more likely to result from the culmination of a growing awareness of people's minds from infancy. This suggests that an impaired theory of mind in autism may be associated with very early developmental difficulties.

10.1 Reading Emotions

Clinically normal children show some ability to distinguish other people's basic emotions around 2-4 months (Field et al., 1982), matching pictures of faces to sounds of emotions around 7 months (Walker, 1982). Autistic children are able to name basic emotions and even predict whether story characters will be happy/sad based on the presence or absence of a desired goal, as long as no understanding of false belief is introduced to the task (Capps et al., 1992). It seems that autistic children have the necessary vocabulary of emotion words but differ form matched controls in their ability to use it (Hobson, 1989).

Children with autism appear to see faces as a pattern that must be interpreted in order to name an expression rather than the automatic experience of knowing how someone else feels from their expression. This was demonstrated by Hobson et al.'s (1988) presentation of faces right side up and upside down. Typically developing children were
much poorer at naming faces emotion upside down, while children with autism were almost as good at naming upside down images as they were at right side up. It seemed that autistic children used a single cognitive strategy, while typically developing children used different strategies for facial stimuli and apparently more random patterns. In typical development, information processing seems primed for emotional meaning, possibly processed via a rapid affective channel as opposed to a more laborious cognitive problem solving route.

10.2 Joint Attention and Pretend Play

Before children learn to speak they develop ways of communicating with those around them, which indicates a growing awareness of other people’s minds even in infancy. Children develop ‘triadic representations’ in their minds, of another person looking at an object which he/she can also see (Baron-Cohen, 1995). In addition to simply monitoring the direction of another person’s eyes, there is evidence to suggest an awareness of the person’s attitude about what he/she sees (Hobson, 1993). Infants crawling towards a visual cliff tended to refer to their mothers’ expression to decide on their course of action; to go on if she looked relaxed, or turn back/freeze if she looked fearful (Sorce and Emde, 1981). Similarly, typically developing children attend to their mother’s attitude towards new toys before deciding to play with them or not (Hornik et al., 1987).

Children with autism differ in their use of joint attention behaviours, including gaze monitoring and pointing, which typically developing children usually achieve around 9
months (Baron-Cohen et al., 1992). Autistic children have been thought to match typically developing children in their use of imperative behaviours (associated with a request for help), but are impaired in their use of declarative behaviours (to share something of interest) (Mundy et al., 1994). Phillips et al. (1995) demonstrated that non-autistic children gained help from the experimenter when a desired object was out of reach by making eye contact with him; and then shifting their gaze between his eyes and the object. Children with autism either ignored him and tried to climb on some furniture to reach for themselves, or dragged him to the location and pushed up his arm in a mechanical way.

The distinction between impaired and preserved joint attention skills is associated with the extent to which children perceive the other person's state of mind (Charman, 1997). Autistic children do not necessarily make eye contact with adults less frequently but do seem to use eye contact less to communicate with another person, or combine eye contact with and affective exchanges, apparently less motivated to achieve an interpersonal connection with others (Charman, 1998). Infants' distress at interacting with their primary carers via a slightly delayed video link-up illustrates their sensitivity to attunement with significant others (Murray and Trevarthen, 1985). Hobson (1993) proposed that this drive towards a co-ordinated exchange, rewarded by the arousal associated with a emotional connection, could be a central impairment in autism.

In typical development the emergence of pretend play follows the acquisition of joint
attention behaviours, at around 14 months (Leslie, 1987). Autistic children’s play seems specifically impaired in symbolic pretence, while functional activities, such as skipping, are less impaired (Baron-Cohen, 1987). The areas of difficulty in pretence are again those associated with the expression of an individual’s personal attitude towards objects.

Together, joint attention skills and pretend play account for over 60% of the variance in children’s performance on language scales (Charman, 1998a). The relationship between these skills is evident for typically developing children (Bates et al, 1979; Tomasello and Farrar, 1986) and for children with autism (Mundy et al., 1994). It might be that children’s use of joint attention to learn new linguistic labels (Tomasello, 1995) is an opportunity missed by autistic children who rarely engage in this form of interpersonal activity (Sigman, 1998).

More than this, however, it seems that children with autism have a specific difficulty with using language to communicate rather than simply as a labelling code (Happé, 1994c). If children with autism are not primed to interpret information for interpersonal relevance, they would not be expected to use language as a communicative tool (Hobson, 1993; Frith and Happé 1994). Leslie (1987) proposed that engagement in pretend play was a precursor for theory of mind ability because of the child’s ability to form metarepresentations by decoupling reality from pretence. The association between these skills might be more a reflection of a shared understanding of the participants’ intent, with much earlier origins which seem specifically impaired in autism (Charman, 1997).
Baron-Cohen (1995) suggested that children theorised about others' states of minds after they had become aware of mental states in themselves. He suggested that this occurred from observing the similarity between their own bodies and other people’s, and arriving at the conclusion that others must also have mental experiences like their own. The ‘theorising’ is proposed to arise from observations of other people’s expressions and behaviour, and interpreting them on the basis of desires and goals and their own experiences.

Hobson (1993) disagreed with the view that internal states are unobservable and that the child has to develop an hypothesis about the relationship between minds and bodies in order become aware that people have minds. Instead he suggested that through “episodes of intersubjectively co-ordinated experience .. an infant eventually becomes aware that people have minds” (Hobson, 1993, p.117). He described how people describe someone’s expression automatically by the emotion it evokes: “He was smiling,” rather than describing the physical appearance and then theorising about the most appropriate emotion to match with it (Hobson, 1993). He suggested that this was a “biologically-given...mechanism for establishing connectedness of mind between infant and care-giver” (Hobson, 1993, p.117) which stimulates the child’s awareness that bodies have minds.
Hobson argued that an impairment in this mechanism accounted for the characteristics associated with autism. In turn, he suggested that this impairment would have implications for the subsequent development of skills which are dependent on an awareness of other people's intentions, beliefs and emotions. This would account for the apparent distinction between skills associated with an 'I-You' understanding, and those dependent on 'I-It' processing (Hobson, 1993).

It has been proposed that greater opportunities to learn about mental states from more varied and increased opportunities to learn about mental states in larger families, is associated with faster theory of mind acquisition (Perner et al., 1994). Interestingly, the effect of family size seems particularly strong for children with poorer language skills, suggesting that a wider opportunity for social interaction at home can compensate for the effects of language difficulties in the development of theory of mind (Jenkins and Astington, 1996).

Research has found that congenitally blind (Minter et al., 1998) and deaf (Peterson and Siegal, 1995) children perform on theory of mind tasks in line with autistic children. This might result from reduced opportunities for learning about other people's mental states through pretend play and social interaction (Dunn, 1996) in families who are unable to communicate about mental states in ways accessible to sight/hearing impaired children (Russell et al., 1998). However, these children experience age related delays, rather than life-long disorder of mentalising ability, with an improvement in theory of mind ability
after 13 years (Russell et al., 1998).

There has been an emphasis in this paper on the utility of additional rather than mutually exclusive accounts of autism. While cognitive, linguistic and affective theories can offer complementary accounts of the preserved and impaired skills associated with autism, debate continues regarding the primary underlying impairment. A useful approach would emphasise the interacting nature of development across the domains over time, and in relation to associated genetic and neurobiological perspectives (Bailey et al, 1996).
12 Rationale for this Study

12.1 Advanced Theory of mind Measures

The theory of mind account of autism evolved from the finding that only 20% of autistic children were able to identify story characters’ false belief compared to at least 80% of typically developing and Down syndrome children. Replications of theory of mind studies with various control groups supported an autism specific deficit in the ability to form mental representations for the majority of participants.

There was, however, a persistent group of children with autism who displayed characteristics across the triad but who were able to pass theory of mind tasks. If autistic individuals with obvious social and communication impairments pass theory of mind tasks, the explanatory power of the hypothesis is weakened. This assumes, however, that autistic children passing theory of mind tasks have equivalent mentalising abilities to typically developing controls from whom their performance scores do not differ significantly.

The proponents of the theory of mind hypothesis emphasised that tasks had originally been designed for typically developing children aged 3 to 5 years, suitable for the autistic children with verbal age within this range who participated in the early research. As more recent studies have assessed the abilities of autistic children with higher verbal ages, with
less language delay, it should not be surprising that there is an increasing tendency for autistic participants to perform at ceiling.

This study was carried out in order to investigate the utility of what will be referred to as ‘advanced’ theory of mind tasks in distinguishing between high functioning autistic individuals and typically developing controls. Happé’s (1994b) Strange Stories were devised for more able children with autism who pass what could be termed ‘standard’ false belief tasks. The group of high functioning autistic individuals who passed standard false belief tasks in Happé’s study (1994a), performed similarly to typically developing children, but differed from adults of a similar chronological age. However, only tentative conclusions could be made about the similarity or difference in performance as groups were not matched on verbal or intellectual ability.

The correlation between verbal ability and theory of mind performance on the Strange Stories raised the question whether the increased challenge of a relatively complex story comprehension and explanatory task accounted for difficulties experienced by the autistic group, for whom language impairments are a defining characteristic. In an attempt to vary the language content of the tasks, Happé (unpublished) went on to devise a set of mentalising and physical cartoons.

Baron-Cohen et al. (1997) also developed an advanced task initially designed for autistic adults in which they were asked to choose one of two target words to describe
photographs of people’s eyes. An adaptation of the task has been developed for children (Baron-Cohen, unpublished) in which they must select one of four emotion or cognition words to describe the eyes.

This study used an updated version of the Strange Stories (Happé, unpublished). Previous work has compared some characteristics of participants but few have matched groups for chronological age, verbal age and IQ. Recognition that these variables may be associated with non-mentalising aspects of task performance suggests that they should be controlled for in a comparison of theory of mind abilities between autistic and matched typically developing children.

The majority of experimental group matching makes the assumption that typically developing children have equivalent chronological and verbal ages. However, unexpected performance of typically developing children in some studies queries this. In this study, particular care was be taken in matching groups for verbal age by assessing typically developing as well as autistic participants on the language and IQ measures. It would be of particular interest to identify whether there were characteristics associated with higher performance on theory of mind tasks.

Most studies which have measured verbal ability have used a vocabulary test, either the Peabody Picture Vocabulary Test (PPVT, Dunn, 1965) or the British equivalent test, the British Picture Vocabulary Test (BPVS, Dunn et al.,1982). Whilst there is a benefit from
studies using the same verbal ability measure in order to aid comparisons across studies, this is a relatively restricted measure of verbal ability. With increasing interest in the relationship between verbal ability and theory of mind task performance, there would be a benefit of taking a more detailed measure of language ability. Consequently, this study used four separate sub-tests from the CELF-R (Semel et al., 1987), including both expressive and receptive language measures in the calculation of verbal age.

12.2 Is Autism associated with a mentalising deficit or impairment?

The false belief tasks designed to measure theory of mind abilities are based on a pass/fail model of deficit. Initially it was thought that theory of mind was an all or none ability, achieved by typically developing children by the age of 4 years, but absent in autism. Interest in the typical development of theory of mind stimulated research into the effect of small modifications to the original tasks. It seemed that providing a concrete representation of the false belief, or changing the wording of the task, increased the proportion of children able to pass tasks at younger ages. This research influenced the work into the mentalising abilities of autistic children, which also suggested that amendments to tasks resulted in improved performance.

There was increasing pressure for an explanation that might account for some children with autism being able to pass false belief tasks. It was suggested that children with autism might develop compensatory strategies to work out appropriate answers to the tasks without depending on an innate understanding of people’s thoughts and feelings.
Further research both supported this hypothesis and also provided additional explanations, including the suggestion that some autistic individuals may in fact develop some mentalising abilities.

This challenged the argument for an absence of mentalising ability and raised an alternative view of a continuum of skills. Research into early precursors of theory of mind in typical development proposed that the ability of children to pass false belief tasks at 4 years of age was the culmination of a developing awareness of the mindfulness of bodies from the early months. This is in contrast to the view that there is a sudden radical shift in the ability to form meta-representations, achieved by the automatic 'switching on' of an innate decoupling mechanism.

Some children with autism, particularly those with IQ within the normal range with a verbal age over 6 years, are able to pass standard theory of mind tasks. It remains unclear how their mentalising abilities compare to typically developing children of the same age and ability. Studies using theory of mind tasks now seem to highlight the variety of performance across tasks and the limitations pass/fail research designs. This study, therefore, intends to address the lack of work on mentalising abilities of autistic children across a scale of performance.

The study will investigate performance on currently unpublished advanced theory of mind tasks that are scored on performance scales rather than pass/fail data. The original
Strange Story task was scored according to correct/incorrect explanations. The version used in the current study uses a scaled scoring system, giving credit for partially correct answers for each of the 13 items. The Cartoons task was scored on a performance scale of 0-3 for 20 pictures, in order to explore the variation in responses for both physical and mentalising items. The Eyes task was composed of pass/fail items, but produced an overall score on a scale of 0-27. The unpublished nature of these tasks meant that great care was taken in this study to ensure reliability of scoring using detailed scoring criteria and the double marking of responses for the Stories and Cartoons to enable inter-rater reliability checks.

In addition this study assessed the level and type of mentalising language used. It might be that mentalising abilities range according to the frequency of use, but also the complexity and domain. This study differed from most previous studies in the area of theory of mind tasks by scoring children’s responses for the number of statements with a mentalising explanation but also for the level of mental representation (first, second, or third order) and the type of language (desire, emotion, or cognitive).
13  Research Questions

1. Do advanced theory of mind tasks provide developmentally appropriate measures for able autistic 6-12 year olds, by discriminating performance from matched typically developing controls in a way that standard theory of mind tasks do not?

2. Are the difficulties in mentalising associated with autism more usefully described as a continuum of impairments, rather than an absence of a theory of mind?

3. Are there characteristics that are associated with mentalising ability in autism?

4. Is there evidence to support a distinction between ‘cognitive’ and ‘affective’ aspects of mentalising abilities?
Method

1 Participants

Two groups were assessed in this study: a group of children diagnosed as falling at the higher functioning end of the autistic spectrum, and a group of non-autistic children matched for verbal age. All the children in the study had English as a first language.

1.1 Autism Group

The autism group was composed of children between the ages of 6-12 years who were diagnosed with an autistic spectrum disorder according to ICD-10 criteria for Childhood Autism, Asperger Syndrome, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS). In order to investigate characteristics associated with autism, by minimising the confounding effects of additional learning disabilities, children were recruited who had IQs within the normal range.

Only 25% of autistic children have preserved cognitive ability, and do not fit the classic presentation of profound disabilities across the triad of impairments, who are increasingly diagnosed in secondary child development teams. Consequently, for this study a tertiary child assessment centre was contacted, that was known to receive referrals from primary and secondary services requesting detailed assessments for children requiring differential diagnoses of higher functioning children.
Parents of the thirty-five children most recently assessed and diagnosed at the centre with an autism spectrum disorder and who had an IQ of 70 or above, were invited to join the study. Of these twenty-five provisionally agreed to get involved and twenty children completed the assessment.

The autistic group, composed of 18 boys and 2 girls, had a mean group IQ of 99.8 (SD 20.6) in the range of 71 to 159, measured on the Kaufman-Assessment Battery for Children (K-ABC; Kaufman, 1983) or Wechsler Intelligence Scales for Children III (WISC III; Wechsler, 1991). The mean chronological age of the group was 9 years 4 months (SD 1.6), which ranged from 6 years 5 months and 12 years 1 month. The mean verbal age measured on the Clinical Evaluation of Language Fundamentals-Revised (CELF-R UK; Semel, Wiig and Secord, 1987) was 8 years 3 months (SD 2.4), from 6 years 0 months to 15 years 0 months.

1.2 Typically Developing Group

Twenty non-autistic children were recruited from North London mainstream schools. Headteachers were asked to identify children between 6-12 years performing within the average range, who did not have additional social or communication difficulties, and invite them to join the study. Of the parents who agreed to participate, children were recruited according to group matching with the autistic participants as closely as possible.
The group mean chronological age was 8 years 8 months (SD 1.9), falling in the range of 6 years 4 months to 12 years 5 months. These typically developing children had mean verbal age on the CELF-R of 8 years 10 months (SD 2.2) falling between 5 years 9 months and 13 years 1 month. The group mean IQ was 106.0 (SD 11.6) ranging from 75 to 127 on the WISC. The group was composed of 5 boys and 15 girls.

2   Presentation

Children were assessed in a single session at the assessment centre, at school, or in their homes. Participants were assessed individually in a quiet room by the same experimenter. At the beginning of the assessment each child was presented with a chart of blank boxes representing the number of activities to be completed before a break/end of the session.

Materials were presented to each child by the experimenter who sat beside the child at a desk facing a blank wall. This minimised visual distractions, and any social anxiety that might have arisen from facing the experimenter. Each task had a clear beginning and end, which were made explicit to the child, with verbal prompts when the child was half way through the task. This structuring of the session and the physical environment was based on practices recommended in the TEACCH psycho-educational programme (Schopler & Mesibov, 1987).
The tasks were presented in a forward order: Strange Stories, Sally Ann, Cartoons, Picture Sequencing, Eyes and then Smarties, for half of each group, and in reverse for the other half of each group to counterbalance the effects of fatigue. Some children had completed the CELF-R and/or K-ABC/WISC within the previous 18 months as part of their initial assessment at the children’s centre. Participants who needed the CELF-R and/or WISC as part of this assessment received them after the theory of mind tests had been administered.

3 Measures

3.1 Participant Profile

3.1.1 Language

Children’s language ability was assessed on the Clinical Evaluation of Language Fundamentals-Revised (CELF-R UK; Semel et al., 1987). The assessment provides standardised scores for receptive language, expressive language and a total language score, with mean of 100. The standardised scores were calculated from two out of three sub-tests on the receptive scale; Oral Directions and Semantic Relationships, and the expressive scale; Formulated Sentences and Sentence Assembly.

3.1.2 IQ

The Wechsler Intelligence Scales for Children–III (WISC-III R, Wechsler, 1991) is a standardised measure of general intellectual functioning, with established reliability and
validity data. Participants in the typically developing group and the majority of the autism group were tested on a short form of the WISC-III, composed of the Similarities, Arithmetic, Picture Completion, and Block Design sub-tests. This short version was considered to be the best compromise between psychometric, clinical and practical factors (Kaufman et al., 1996).

Some children in the autism group had been tested on the Kaufman Assessment Battery for Children (K-ABC, Kaufman, 1983) within the previous year at the assessment centre, and these scores were used in the research. The K-ABC has been demonstrated as a measure of general intellectual functioning unaffected by cognitive styles associated with autistic children (Stavrou & French, 1992). The literature supports a good correlation between the WISC-III and the K-ABC of $r=0.68$ (Rust & Yates, 1997).

3.2 STANDARD THEORY OF MIND MEASURES

The measures were grouped into what will be referred to as ‘standard’ theory of mind measures and ‘advanced’ measures. The standard measures were the Sally-Ann (Baron-Cohen et al., 1985) and Smarties (Perner et al.,1989) false belief tasks, and the Picture Sequencing task (Baron-Cohen et al., 1986). Advanced measures were the Strange Stories (Happé, 1994; Happé, unpublished), Cartoons (Happé, unpublished), and Eyes task (Baron-Cohen et al., 1997; Baron-Cohen, unpublished). See appendix for samples of items and scoring criteria.
Sally Ann Task

Design

The Sally-Anne task (Baron-Cohen et al., 1985) required children to predict where a doll would look for her marble which had in fact been moved to a new location by another character without her 'seeing.'

Procedure

Children were introduced to each of the two dolls, and then asked to name each one to check their memory of who was who. The child was shown Sally putting a large button into her box. The Sally doll then left the scene. Ann then moved the button from the box into her bag. Sally then returned, and the experimenter asked the child a belief question, "Where will Sally look for her button?" a question about reality, "Where is the button really?" and a memory question, "Where was the button at the beginning?"

Scoring

Each child was given 1 for a correct or 0 for an incorrect answer for the belief question, reality question, and the memory question. The reality and memory questions were designed to ascertain whether mistakes on the belief question were specifically associated with mentalising difficulties, or because of general cognitive difficulties associated with memory and language delay.
**Smarties Task**

**Design**

The Smarties task (Perner et al., 1989) required children to predict what another person would think was in a tube of sweets after being shown the true contents.

**Procedure**

The participants were presented with a closed Smarties tube and asked, “Look what I have got here, what do you think is inside?” After the child replied the experimenter opened the tube to reveal the true contents and said, “Well actually, there is a pencil” and then re-closed the tube. After a pause the child was asked a question about another person’s false belief. “Your Mum (or whoever had brought the child to the assessment) has not seen me show you this tube, but if I asked her to come in here, and I showed her this tube, what would she think was inside?” A reality question was also asked, “What is really in here?” and a memory question, “What did you think was in here when I first asked you?” All answers were recorded at the end of the task.

**Scoring**

Children scored 1 for a correct response for predicting that the adult would think there were Smarties inside, and 0 for an incorrect response. The reality and memory questions were control items, intended to ensure whether any mistakes were the result of mentalising difficulties rather than problems understanding the task.
Picture Sequencing Task

Design

Children were presented with 15 sets of picture cards, which could be arranged to illustrate a story (Baron-Cohen et al., 1986). Each story was made up of 4 brightly coloured, laminated cards. The test did not include written language in the presentation nor in the child's response, although participants needed to understand the presenter's verbal explanation of the activity.

Three story conditions were used in this study, as recommended by Baron-Cohen et al. (1985). 'Mechanical' stories involved people and/or objects, 'Behavioural' stories involved people in non-mentalising social situations, and 'Intentional' stories included people interacting in a way that was motivated by their mental states.

Procedure

Participants were told that there were some sets of pictures to look at. The first set of four cards was put on the table in front of the child in a jumbled up order, but all the right way up. The experimenter said that each set of pictures made a story. The child's attention was drawn to a velcro board in front of him/her, with a space for each card. "I will put the first card in place for each story, then I would like you to choose the next card that continues the story, and carry on until all the cards fill the board and the story is
Scoring

The first card was put in position by the experimenter. Each story sequence was marked out a possible 2 for the remaining cards. According to the original study, 2 points were given for a correct order, and 1 point was given for the last card being in the right place. The 9 stories were categorised as 3 intentional, 3 behavioural and 3 physical, with composite scores out of a possible 6 for each story type.

3.3 Advanced Theory of Mind Measures

Strange Stories

Design

The Strange Stories (Happe, 1994) were developed as an advanced measure of theory of mind, for children able to pass false belief tasks. An updated version of these stories was used in this study (Happe, unpublished). Children were presented with 5 stories based on physical/mechanical events, and 8 stories that required an understanding of characters' thoughts. This task had requirements of receptive language for understanding the story, and expressive language to explain reasons for characters' behaviour.

Procedure

The experimenter began the task by showing the child a booklet and saying, "On each
page there is a short story which I will read to you. I will then read you the question on the next page, which I would like you to answer." A practice story was read out first, followed by the physical and mentalising stories in a randomised order. The stories were read aloud for all children, even if they could read for themselves, in an attempt to control for the type of verbal skills used in the task. The stimulus question was asked with the story and picture still in view to minimise the loading of memory skills. Positive comments were made throughout the tests but no feedback was given about accuracy of responses, and no prompts were given. If children corrected themselves they were given credit for their best answer. Responses to each question were written in full at the time of testing.

Scoring

If a child gave more than one answer only the most appropriate was marked. If s/he gave both mentalising and physical explanations, the mentalising response was scored. Omissions were marked as errors. Intraclass correlation (ICC; Shrout and Fleiss Model 3) reliability estimates for 95% confidence interval, were calculated for the 0-2 complexity scoring. 12 out of 13 stories had a reliability measure above 0.83. The range of estimates for the 13 stories was 0.77-1.00, with mean of 0.95. This reflects a very high inter-rater reliability and indicates good reliability for the complexity score on the story ratings. ICC reliability estimates were also calculated for the 0-3 level of mentalising language on the 8 mentalising stories. 5 out of 8 mentalising stories had reliability above 0.81, with a range of 0.53-0.95. The mean ICC estimate at CI 95% was 0.79. This demonstrates a good
inter-rater reliability for scoring levels of mentalising language.

The 13 stories were rated on a 0-2 scale as fully correct, being partially correct or incorrect according to criteria. For example a 2 point answer might be, *The prisoner fooled the enemy by telling them the true location of his army, as then they were sure to look somewhere else.* A 1 point version for the same story would demonstrate a partial understanding, *The prisoner told the enemy where his army really was.* And a 0 point answer would be incorrect, such as *The prisoner did not tell the enemy anything about his tanks.*

The 8 mentalising stories were also marked for the use of mental state language, including words such as *thinks, knows, wants, pretends, lies, loves, scared, polite.* The mentalising total was the number of items for which any type of mental state explanation was given, that is a score out of a possible 8. These mentalising statements were then rated for the level of mental representation the child used. A first order representation (*He lied*) scored 1, second order scored 2 (*They will think that he lied*), and third order mentalising statement (*He knew that they would think that he would lie*) scored 3. A cumulative mentalising score was calculated by computing the 0-3 scores for each response, out of a possible 24. A score for the number of incorrectly attributed mental states was calculated for the 8 mental state stories (e.g. *He wanted to tell where his army really was*).
Cartoons

Design

Participants were also asked to answer the question, “Why is this funny?” for twenty black and white cartoons. Half of them required an understanding of characters' mental states, while the others were based on mechanical/behavioural humour (Happe, unpublished). This test had minimal language in the presentation but was dependent on the child to understand the task, particularly the concept of ‘funny,’ and to describe what s/he understood by the picture.

Procedure

Each child was presented with the first cartoon, and was told, “I have some pictures to show you. In each one there is something funny happening. I would like you to tell me why it is funny.” No prompts were given to participants, but positive encouragement was given throughout.

Scoring

Children were rated on a 0-3 scale according to the appropriateness of their explanation (see appendix for criteria). For example, a 3 point answer would demonstrate an understanding of mental states where relevant. For example, The bull did not know that behind the cape, the matador was hiding a weight that would knock him out. A 2 point answer would demonstrate some understanding of what was implied in the cartoon, The bull will get knocked out by the weight, and a 1 point answer would simply describe the
picture, There is a heavy weight behind the cape. 0 points would be given to the mention of an incorrect or irrelevant aspect of the cartoon, The man does not have any trousers on. The physical cartoons were also scored 0-3 for complexity of answers. For example, The fisherman has been sitting for so long that a spider has spun its web on the line, would score 3 points; The spider has spun its web into the line, would score 2; a 1 point answer may be that There is a fisherman waiting to catch a fish; while a comment about his appearance would score 0.

Children who gave more than one response were given credit for their best response, and for any explanations which included mentalising statements. Omissions scored 0 points. An intraclass correlation estimate was calculated for the 0-3 scoring on the 20 cartoons. The estimates ranged from 0.62-1.00, with 16 out of 20 cartoons scoring above 0.84. The mean ICC estimate was 0.91. This reflects a very high level of inter-rater reliability for the mentalising and physical cartoons.

Eyes Task

Design

Participants were presented with a child version of the Adult's Eyes Test (Baron-Cohen et al., 1997; Baron-Cohen, unpublished). This comprised of 27 black and white photographs of men and women's eyes from magazines, from above the brow line to halfway down the nose. Each photograph was surrounded by four words describing potential feelings, from which the child had to choose the best match (Baron-Cohen et al.,
in press). Each word was pointed to and read out to the child in turn.

Procedure

Each child was presented with a booklet of 28 photocopied photographs and was told, "I would like you to look at these pictures I have of people's eyes." The child was presented with the practice item. "Look at this one. Around the photograph are four words that I will read out to you. I would like you to tell me which word best describes the way the person is thinking or feeling." The experimenter then pointed to and read each word in turn, and encouraged the child to choose the best one. After the practice the child was told, "I have some more for you to try, remember just choose the best word for how the person is thinking or feeling."

Scoring

The responses were scored as 1 for correct or 0 incorrect answers, according to the original study (Baron-Cohen, unpublished).
Results

1 Group Differences

1.1 Participant Characteristics

1.1.1 Chronological Age

Table 1 shows the means of participants' characteristics across the autistic and typically developing groups, with standard deviations and the range of scores. The age of the children in the study was between 6 and 12 years. The mean age of the autistic spectrum participants (n=20) was 9 years 4 months (SD 1.6), and 8 years 8 months (SD 1.9) for the typically developing group (n=20). This did not represent a significant difference in chronological age between the two groups.

1.1.2 Verbal Age

Comparing the verbal age, measured by the CELF-R (Semel et al., 1987), the mean for the autistic group (n=20) was 8 years 3 months (SD 2.4) which did not differ significantly from the typically developing group (n=20) with mean age of 8 years 10 months (SD 2.2). The range in verbal age in the autistic group was from 5 years 0 months to 15 years 0 months and for the typically developing group from 6 years 3 months to 12 years 7 months.
Table 1
Participant Characteristics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Autism Group (n=20)</th>
<th>Typical Group (n=20)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Age¹</td>
<td>9.4</td>
<td>1.6</td>
<td>6.5-12.1</td>
</tr>
<tr>
<td>Verbal Age²</td>
<td>8.3</td>
<td>2.4</td>
<td>5.0-15.0</td>
</tr>
<tr>
<td>Express. Lang³</td>
<td>83.4</td>
<td>13.1</td>
<td>59-109</td>
</tr>
<tr>
<td>Recept. Lang³</td>
<td>88.8</td>
<td>20.5</td>
<td>59-130</td>
</tr>
<tr>
<td>Total Score³</td>
<td>86.7</td>
<td>19.2</td>
<td>59-133</td>
</tr>
<tr>
<td>IQ³</td>
<td>99.8</td>
<td>20.6</td>
<td>71-159</td>
</tr>
</tbody>
</table>

SD = Standard Deviation

*p < 0.05; **p < 0.01; ***p < 0.005.

1 Chronological age in years and months
2 Verbal age equivalent in years and months calculated from CELF-R
3 Standardised language score (mean=100) on the CELF-R
4 Standardised score (mean=100) on either WISC or K-ABC

1.1.3 Standardised Language Scores

An important feature of this research was the careful matching of the typically developing group for chronological age, verbal age and IQ. Language delay is one of the diagnostic criteria for autism, and therefore an expected characteristic of the autistic group. It will take autistic children longer to acquire language skills than a typically developing child matched for verbal age. It was as expected, therefore that mean receptive language scores for the autistic group (mean 88.8; SD 20.5) were lower than for typically developing group (mean 109.7; SD 17.6) to a significant degree (t = -3.45, 38df, p < .005).

Similarly, there was a significant difference for expressive language scores (t = 2.89, 38df, p < .01) between the autistic group (mean 83.4; SD 19.6) and the typically developing group (mean 94.1; SD 10.0). Finally there was a significant difference in the
total language scores ($t = 3.03, \text{ } 38df, p < .005$), between the autistic (mean 86.7; SD 19.2) and typically developing group (101.3; SD 10.0).

### 1.1.4 IQ

There was no significant difference between the group means for intellectual ability, measured by the K-ABC (Kaufman, 1983) or WISC-III (Weschler, 1991). The full scale group mean for autistic children was 99.8 (SD 20.6), and 106.0 (SD 11.6) for the typically developing children.

### 1.1.5 Gender

The two groups were not equally matched for gender in this study, with 18 boys in the autistic group, and 16 girls in the typically developing group. Comparisons were made between the performance of boys and girls within the groups, which showed that any differences in performance between the two groups was not the result of differences in the gender ratio. The only variable which showed significant gender differences in the autistic group was for performance on the physical cartoons ($t = 2.43, \text{ } 18df, p < .05$), for which the boys performed better than the girls. There were no significant differences in performance between boys and girls in the typically developing group.

### 1.2 Standard Measures

#### 1.2.1 False Belief Tasks

Table 2 shows the performance of autistic and typically developing children on the Sally-
Ann and Smarties belief, reality and memory questions. All the children in the study answered the reality and memory questions correctly for both false belief tasks. As expected, the majority of the typically developing group passed the belief questions for the Sally-Ann (95%) and they all passed the Smarties task. The autistic group also performed well, 80% passed the Sally Ann belief question and 95% passed the Smarties belief question. Chi-square analyses found no significant differences between the performance of the two groups for either the Sally-Ann task $\chi^2(1, N=20) = 2.06, p >.05$, or the Smarties task $\chi^2(1, N=20) = 1.03, p >.05$. 

Table 2
Group Performance on Standard Theory of Mind Tasks

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Autism Group (n=20)</th>
<th>Typically Developing (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td>Sally-Ann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief</td>
<td>16 (80%)</td>
<td>4 (20%)</td>
</tr>
<tr>
<td>Reality</td>
<td>20 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Memory</td>
<td>20 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Smarties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief</td>
<td>19 (95%)</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>Reality</td>
<td>20 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Memory</td>
<td>20 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Picture Sequence $^1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional</td>
<td>4.75</td>
<td>1.62</td>
</tr>
<tr>
<td>Behavioural</td>
<td>5.40</td>
<td>1.14</td>
</tr>
<tr>
<td>Physical</td>
<td>5.70</td>
<td>0.73</td>
</tr>
</tbody>
</table>

SD = Standard Deviation

1 Picture Sequence, score: 0-6
1.2.2  Picture Sequence Task

There were no significant differences between the performance of the autistic group and
the typically developing group for the physical ($t = 0.76$, $38df$, $p > .05$), behavioural ($t =
0.66$, $38df$, $p > .05$) or intentional ($t = 0.68$, $38df$, $p > .05$) story types.

1.3  'ADVANCED' MEASURES

1.2.1  Strange Stories

Table 3 illustrates the scores for the tasks for each group, and any differences between
them. The mean score for the autistic group on the physical stories was 6.0 (SD 2.5), and
6.7 (1.5) for the typically developing group. On the mentalising stories, the mean score
for the autistic group was 7.7 (SD 4.9), compared to the typically developing group’s
mean score of 12.0 (SD 3.2). This represents a significant difference between the groups’
performance on the complexity score for the mentalising stories ($t = 3.34$, $38df$, $p < .005$), but not for the physical stories ($t = 1.17$, $38df$, $p > .05$).

Story explanations were rated for the number and type (first, second or third order) of
mentalising statements that were used. The mentalising total score indicated the number
of mentalising items in which any mental state language was used. The autistic group had
a mean score of 6.1 (SD 1.6), and the typically developing group had a mean score of 7.2
(SD 0.8) out of a possible 8. Despite the high performance of both groups, this did
represent a significant difference between the use of a mental state explanations for the

87
mentalising stories between the two groups (t = 2.70, 38df, p < .05). The autistic group's cumulative mentalising score for the level of mental state language used was significantly lower than for the typically developing group (t = 3.14, 30df, p < .005). In addition, the autistic group gave significantly more incorrect mental state attributions than the typically developing group (t = 2.38, 27df, p < .05).

In particular, t-tests identified the source of the difference in the level of mentalising statements used by the two groups. The autistic group made fewer second order mental state terms (mean 0.8; SD 1.2) than the typically developing group (mean 1.6; SD 1.3) to a statistically significant degree (t = 2.04, 38df, p < .05). There were no significant differences in the number of first order mentalising statements between the autistic (mean 5.5; SD 1.5) and typically developing (mean 5.8; SD 1.8) groups (t = 0.57, 38df, p > .05). Similarly, the number of third order statements made by autistic (mean 0.1; SD 0.2) and typically developing children (mean 0.1; SD 0.3) showed no significant differences (t = 0.59, 38df, p > .05).

Responses were categorised according to desire statements (such as want, like, love), cognitive statements (including think, know, lie, pretend), and emotion words (including scared, angry, polite, upset). Each story response was scored for every mental state term used so the range of scores was greater than the total number of stories. There were no significant differences in the number of desire words between the autistic (mean 3.3; SD 2.0) and typically developing (mean 4.3; SD 1.5) groups (t = 1.88, 38df, p > .05). The use of emotion words made by autistic (mean 2.4; SD 1.8) and typically developing
children (mean 2.2; SD 7.2) was similar (t = .30, 38df, p > .05). Finally the use of cognitive words used by the autistic (mean 3.7; SD 2.6) and typically developing (mean 4.0; SD 2.4) groups did not differ (t = .44, 38df, p > .05).

Table 3
Group Performance on Advanced Theory of Mind Tasks

<table>
<thead>
<tr>
<th>TASKS</th>
<th>Autism Group</th>
<th>Typical Group</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean SD range</td>
<td>mean SD range</td>
<td></td>
</tr>
<tr>
<td><strong>Strange Stories</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentalising complexity¹</td>
<td>7.7 4.86 0-15</td>
<td>12.0 3.21 4-16</td>
<td>3.34***</td>
</tr>
<tr>
<td>Physical complexity²</td>
<td>6.0 2.48 0-9</td>
<td>6.7 1.45 4-9</td>
<td>1.17</td>
</tr>
<tr>
<td>Mental state answers³</td>
<td>6.1 1.62 2-8</td>
<td>7.2 0.83 5-8</td>
<td>2.70*</td>
</tr>
<tr>
<td>Mentalising score⁴</td>
<td>7.0 2.48 2-11</td>
<td>9.0 1.40 6-12</td>
<td>3.14***</td>
</tr>
<tr>
<td>Incorrect mental answers⁵</td>
<td>1.3 1.45 0-6</td>
<td>0.4 0.68 0-2</td>
<td>2.38*</td>
</tr>
<tr>
<td><strong>Cartoons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentalising score⁶</td>
<td>15.9 7.20 3-27</td>
<td>17.3 7.38 3-28</td>
<td>0.63</td>
</tr>
<tr>
<td>Physical score⁷</td>
<td>16.6 6.57 5-27</td>
<td>18.1 4.75 11-25</td>
<td>1.01</td>
</tr>
<tr>
<td>Incorrect mental answers⁸</td>
<td>1.7 2.20 0-8</td>
<td>1.4 2.25 0-8</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Eyes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score⁹</td>
<td>14.3 3.77 9-24</td>
<td>17.2 4.23 9-23</td>
<td>2.29*</td>
</tr>
</tbody>
</table>

SD = Standard Deviation * p < 0.05; ** p < 0.01; *** p < 0.005.

¹ Total scores (0-16) for responses to 8 mentalising stories.
² Total scores (0-10) for responses to 5 physical control stories.
³ No. (0-8) of responses in which a mental state term was used in mentalising stories.
⁴ Total score (0-24) for mental representation on mentalising stories.
⁵ No. (0-8) of wrongly attributed mental state terms used in the mentalising stories.
⁶ Total score (0-30) for responses to 10 mentalising cartoons.
⁷ Total score (0-30) for responses to 10 physical control cartoons.
⁸ No. (0-10) of wrong explanations for mentalising cartoons.
⁹ No. (0-27) correct answers on the eyes task.

A post hoc analysis of the different stories was carried out to identify whether some items demonstrated a stronger discriminatory power between the two groups. Figure 1 illustrates the mean complexity scores (0-2) for the 8 mentalising and 5 physical stories, for the autistic and typically developing groups. Owing to the large number of
comparisons carried out on each of the 13 stories, caution was taken by using the 99% confidence interval. Using these criteria, three mentalising stories were identified as discriminating between the two groups. The Armies item required an understanding of double bluff ($t = 3.27$, $38df$, $p < .005$), the Sausages item needed an explanation that the greedy boy was attempting to arouse sympathy in order to get more food ($t = 3.09$, $df 38$, $p < .005$), and an understanding that the Burglar erroneously thought the policeman was aware of his guilt ($t = 3.97$, $38df$, $p < .001$).

In addition to t-test comparisons between scores on the Strange Stories, a repeated measures analysis provided a parsimonious way of comparing the two dependent variables of story type and group performance. The analysis made the assumption that the participants completed two equivalent story sets, with a difference in one set of an additional mentalising component. There are different numbers of the two story types, so a proportion score was calculated for performance on the two story types for the purpose of this analysis.

There was a significant interaction effect between the type of story and the performance of the two groups ($F = 13.62$, $1df$, $p < .005$). This effect is illustrated in Figure 2, which shows that the autistic group performed better on the physical stories than the mentalising stories, while the typically developing group scored higher on the mentalising stories, but performed similarly to the autistic group on physical stories.
Figure 1
Histogram of Mean Complexity Scores for Individual Items on Strange Stories

Mentalising Stories
M1 Brothers
M2 Armies
M3 Sausages
M4 Kittens
M5 Hat
M6 Rabbit
M7 Mrs Peabody
M8 Burglar

Physical Stories
P1 Armies
P2 Jewel Thief
P3 Washing
P4 X-Ray
P5 Storm

Figure 2
Plot of Repeated Measures Analysis of Group Performance by Story Type
1.2.2 Cartoons

The 10 mentalising and 10 physical cartoons were scored on a scale of 0-3 which incorporated both the complexity of participants’ answers and their understanding of the implicit joke in the pictures. Out of a possible score of 30, the autistic group mean was 17 for the physical items and 16 for the mentalising items. These scores did not differ significantly from the typically developing children’s performance of 18 for the physical, and 17 for mentalising cartoons (t = 0.63, 38df, p > .05). The autistic group did not miss the point of the cartoon completely (i.e. score zero) more often than the typically developing group, for either the physical or mentalising items (t = 0.50, 38df, p > .05).

The post hoc item analysis found that the only one cartoon demonstrated discriminatory power at the 95% confidence interval, although this must be taken with caution because of the large number of comparisons carried out. The cartoon was a picture of the bird tugging on a worm which was really part of a monster (t = 2.13, 38df, p < .05). In fact, Figure 3 shows that on some items the autistic group out-performed the typically developing group, although not to a significant degree.

A repeated measures analysis, which again provided a parsimonious comparison of the two dependent variables, cartoon types and the group performance. Both groups found the physical cartoons easier than the mentalising cartoons, as illustrated in Figure 4, with no overall (F = 2.40, 1df, p > .05) or interaction effects (F = 0.10, 1df, p > .05).
Figure 3
Histogram of Mean Scores for Individual Items on Cartoon Task

Mentalising Cartoons
M1 Dog and Hose
M2 Tarzan
M3 Bull Fighter
M4 Monster
M5 Spanner

Physical Cartoons
M6 Snake
M7 Bird Bath
M8 Picnic
M9 Headphones
M10 Hold-UP

M10 M3 M5 M7 M9 P10 P3 P5 P7 P9

Figure 4
Plot of Repeated Measures Analysis of Cartoon Type by Group
1.2.3  Eyes Task

The eyes were scored 0 for incorrect or 1 for correct choices of target words out of 27 items. The mean score for the autistic group was 14 (52 %) and the typically developing group mean was 17 (63 %). There was a significant difference between the performance of the two groups (t = 2.29, 38df, p < .05). Both groups were performing well above chance, which would be 7 out of 27.

Post hoc item analysis identified four photographs that discriminated between the autistic and typically developing group, although only at the 95% confidence interval, which should be considered cautiously because 27 comparisons were calculated on this data. Figure 5 illustrates the unexpectedly finding that the mean score for the autistic group for item 7, interested, was higher than for the typically developing group $\chi^2 (1, N = 40) 5.01, p < .05$. Items 13, thinking about something $\chi^2 (1, N = 40) 5.01, p < .05$, 22 sure about something $\chi^2 (1, N = 40) 4.29, p < .05$ and 26 not believing $\chi^2 (1, N = 40) 5.01, p < .05$, showed some difference between groups in the direction of higher performance by the typically developing group.

2  Performance across Tasks

2.1  Participant Characteristics

Correlations between participant characteristics and task performance are shown in Table 4. For the autistic group, IQ was related to performance on the physical stories ($r = .46, p$
Figure 5
Histogram of Mean Scores for Individual Items on Eyes Task

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Correct Target Word</th>
<th>Item No.</th>
<th>Correct Target Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kind</td>
<td>15</td>
<td>made up her mind</td>
</tr>
<tr>
<td>2</td>
<td>sad</td>
<td>16</td>
<td>a bit worried</td>
</tr>
<tr>
<td>3</td>
<td>friendly</td>
<td>17</td>
<td>thinking about something sad</td>
</tr>
<tr>
<td>4</td>
<td>upset</td>
<td>18</td>
<td>interested</td>
</tr>
<tr>
<td>5</td>
<td>making somebody do something</td>
<td></td>
<td>not pleased</td>
</tr>
<tr>
<td>6</td>
<td>worried</td>
<td>19</td>
<td>interested</td>
</tr>
<tr>
<td>7</td>
<td>interested</td>
<td>20</td>
<td>thinking about something</td>
</tr>
<tr>
<td>8</td>
<td>remembering</td>
<td>21</td>
<td>sure about something</td>
</tr>
<tr>
<td>9</td>
<td>thinking about something</td>
<td>22</td>
<td>serious</td>
</tr>
<tr>
<td>10</td>
<td>hoping</td>
<td>23</td>
<td>worried</td>
</tr>
<tr>
<td>11</td>
<td>serious</td>
<td>24</td>
<td>nervous</td>
</tr>
<tr>
<td>12</td>
<td>thinking about something</td>
<td>25</td>
<td>not believing</td>
</tr>
<tr>
<td>13</td>
<td>thinking about something</td>
<td>26</td>
<td>happy</td>
</tr>
<tr>
<td>14</td>
<td>not believing</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>
Verbal age was correlated with performance on the physical stories \((r = .47, p < .05)\), and performance on the Eyes task \((r = .60, p < .01)\). There was no correlation between task performance and chronological age.

**Table 4**

**Pearson Correlations between Advanced Task Performance and Participant Characteristics**

<table>
<thead>
<tr>
<th>Task</th>
<th>Autistic</th>
<th>Typically Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ment. Complex (^4)</td>
<td>.22</td>
<td>.34</td>
</tr>
<tr>
<td>Phys. Complex (^5)</td>
<td>.30</td>
<td>.47*</td>
</tr>
<tr>
<td>Mentalising (^6)</td>
<td>.05</td>
<td>.24</td>
</tr>
<tr>
<td>Ment. Score (^7)</td>
<td>.01</td>
<td>.20</td>
</tr>
<tr>
<td>Phys. Score (^8)</td>
<td>.06</td>
<td>.22</td>
</tr>
<tr>
<td>Eyes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score (^9)</td>
<td>.13</td>
<td>.60**</td>
</tr>
</tbody>
</table>

\(*p < 0.05; \ **p < 0.01; \ ***p < 0.005.\)

1 Chronological age  
2 Verbal age: CELF-R  
3 IQ: WAIS or K-ABC  
4 Total score (0-16) for responses to 8 mentalising stories  
5 Total score (0-10) for responses to 5 physical control stories  
6 Total score (0-24) for mental representation on mental stories  
7 Total score (0-30) for responses to 10 mentalising cartoons  
8 Total score (0-30) for responses to 10 physical control cartoons  
9 No. (0-27) correct answers on the eyes task

In the typically developing group there was no relationship between IQ and task performance. Verbal age was related to performance on the physical stories \((r = .67, p < .005)\), mentalising stories \((r = .52, p < .05)\) and for the mentalising cartoons \((r = .46, p < .05)\). As expected, chronological and verbal age were strongly correlated in the typically developing group \((r = .89, p < .005)\), and were related to task performance.
2.2 Measures

2.2.1 Strange Stories

Correlations between performance on the different tasks are shown in Table 5. Performance on the mentalising and physical items on the Strange Stories correlated with one another for both autistic (r = .82, p < .005) and typically developing groups (r = .61, p < .005). There was also a relationship between the use of mentalising words by autistic children and their scores on the complexity measures for both the physical (r = .63, p < .01) and mentalising stories (r = .51, p < .05), but only for the mentalising stories in the typically developing group (r = .54, p < .05).

Table 5
Pearson Correlations for Performance across Advanced Theory of Mind Tasks

<table>
<thead>
<tr>
<th></th>
<th>Autistic</th>
<th>Typically Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MC₁ PC² M³ MS⁴ PS⁵</td>
<td>MC PC M MS PS</td>
</tr>
<tr>
<td>Strange Stories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ment. Complex¹</td>
<td>.82****</td>
<td>.61***</td>
</tr>
<tr>
<td>Phys. Complex²</td>
<td>.56* .57**</td>
<td></td>
</tr>
<tr>
<td>Mentalising³</td>
<td>.54* .51* .14</td>
<td>.47* .29 .13 .83****</td>
</tr>
<tr>
<td>Cartoons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ment. Score⁴</td>
<td>.54* .52* .09</td>
<td>.37 .34 -.01</td>
</tr>
<tr>
<td>Phys. Score⁵</td>
<td>.54*</td>
<td>.47* .29 .13 .83****</td>
</tr>
<tr>
<td>Eyes</td>
<td>-.15 .10 -.11</td>
<td>.15 .20 .05 .23 .01</td>
</tr>
<tr>
<td></td>
<td>.51* .57**</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05; ** p< 0.01; ***p<0.005; ****p<0.001.

1 Total score (0-16) for responses to 8 mentalising stories
2 Total score (0-10) for responses to 5 physical control stories
3 Total score (0-24) for mental representation on mental stories
4 Total score (0-30) for responses to 10 mentalising cartoons
5 Total score (0-30) for responses to 10 physical control cartoons
6 No. (0-27) correct answers on the eyes task
2.2.2 Cartoons

Children's performance on the physical cartoons and the mentalising cartoons was highly correlated for both autistic ($r = .94, p < .001$) and typically developing children ($r = .83, p < .005$). In the autistic group, performance on the Stories and cartoons were all interrelated to a similar degree; physical stories with physical ($r = .51, p < .05$) and mentalising cartoons ($r = .52, p < .05$), and the mentalising stories with physical ($r = .54, p < .05$) and mentalising cartoons ($r = .54, p < .05$). For the typically developing group the only significant relationship between these variables was between the mentalising stories and the physical cartoons ($r = .47, p < .05$).

2.2.3 Eyes

A partial correlation controlling for verbal age did not reveal any relationship between the autistic group's performance on the eyes with performance on the other tasks. In the typically developing group, performance on the Eyes task related to performance on both the physical ($r = .57, p < .05$) and mentalising cartoons ($r = .57, p < .01$).

2.3 Controlling for Verbal Age and IQ

In order to control for the impact of language and cognitive abilities on task performance, partial correlations were carried out looking at performance across the measures with chronological age, verbal age and IQ partialled out, which are presented in Table 6.
Table 6
Partial Correlations for Task Performance, controlling for Chronological Age, Verbal Age and IQ.

<table>
<thead>
<tr>
<th></th>
<th>Autistic</th>
<th>Typically Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS¹ PS² M³ MC⁴ PC⁵</td>
<td>MS PS M MC PC</td>
</tr>
<tr>
<td>Strange Stories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ment Complex¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phys Complex²</td>
<td>.78****</td>
<td>.40</td>
</tr>
<tr>
<td>Mentalising³</td>
<td>.53* .57*</td>
<td>.55* .11</td>
</tr>
<tr>
<td>Cartoons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ment. Score⁴</td>
<td>.50* .50* .03</td>
<td>.19 -.07 -.07</td>
</tr>
<tr>
<td>Phys. Score⁵</td>
<td>.50* .48 .94****</td>
<td>.31 -.08 .08 .80****</td>
</tr>
<tr>
<td>Eyes</td>
<td>-.42 -.15 -.36 .06 .13 -.11 .07 -.01 .48* .51*</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05; ** p< 0.01; ***p<0.005; ****p<0.001.

1 Total score (0-16) for responses to 8 mentalising stories
2 Total score (0-10) for responses to 5 physical control stories
3 Total score (0-24) for mental representation on mental stories
4 Total score (0-30) for responses to 10 mentalising cartoons
5 Total score (0-30) for responses to 10 physical control cartoons
6 No. (0-27) correct answers on the eyes task

In the autistic group there was a strong correlation between performance on the two types of Strange Story (r = .78, p < .001). Performance on the two types of Cartoon were also strongly correlated (r = .94, p < .001). The mentalising scores were correlated with both the mentalising (r = .53, p < .05) and physical (r = .57, p < .05) Story scores. The mentalising story scores were correlated with both mentalising (r = .50, p < .05) and physical cartoon scores (r = .50, p < .05). The physical story scores were correlated with the mentalising (r = .50, p < .05) but not the physical cartoon scores (r = .48, p > .05).

There was no significant correlation between the mentalising score for the Strange Stories and either of the Cartoon scores. There was no significant correlation between...
performance on the Eyes and scores for the other tasks.

The typically developing group also showed a strong correlation between the two types of Cartoon \( (r = .80, p < .001) \), but not for the two Strange Stories \( (r = .40, p > .05) \). The mentalising score was correlated with the mentalising story score \( (r = .55, p < .05) \) but not the physical story score \( (r = .11, p < .05) \). There was no significant correlation between the Story scores and the Cartoon scores, nor was there a significant correlation between the mentalising score for the Strange Stories and either of the Cartoon scores. There was a significant correlation between the performance on the Eyes task and both the mentalising \( (r = .48, p < .05) \) and physical cartoons \( (r = .51, p < .05) \).

2.4 **Overall View of Advanced Measures**

In order to identify which measures distinguished the two groups’ performance apart, data were converted into Z scores. This transformation preserves the relative pattern of the distribution but generated scores with an overall mean of 0 and standard deviation of 1, to enable a direct comparison of performance across the different measures. Figure 6 shows the largest group difference was for the mentalising stories, followed by the cumulative mentalising score and then the Eyes score.
Figure 6
Z score comparison of group differences in performance across the advanced tasks

Z Scores for Measures:

ZSTORP Physical Stories  ZSTORM Mentalising Stories
ZSTORRCUM Cumulative Mentalising Score ZEYES Eyes Task
ZCARTP Physcial Cartoons ZCARTM Mentalising Cartoons

clinical group
Table 7
Discriminant Function Analysis of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story mentalising complexity</td>
<td>.52</td>
</tr>
<tr>
<td>Story cumulative mentalising</td>
<td>.49</td>
</tr>
<tr>
<td>Eyes</td>
<td>.36</td>
</tr>
<tr>
<td>Chronological age</td>
<td>-.20</td>
</tr>
<tr>
<td>Story physical complexity</td>
<td>.18</td>
</tr>
<tr>
<td>IQ</td>
<td>.18</td>
</tr>
<tr>
<td>Physical cartoons</td>
<td>.16</td>
</tr>
<tr>
<td>Verbal age</td>
<td>.13</td>
</tr>
<tr>
<td>Mentalising cartoons</td>
<td>.10</td>
</tr>
</tbody>
</table>

Finally, a discriminant analysis was carried out to examine the combined strength of variables to predict group membership. Entering the variables listed in Table 7, there was an overall accuracy in classification at 85%. The multivariate Wilks was highly significant (lambda .48, p < .005). Performance on the mentalising stories was most strongly correlated with the overall function, followed by the cumulative score for mental state language, and then performance on the Eyes task. The groups were matched for age, language ability and IQ so it was not surprising that these variables were not strongly associated with the function.
Discussion

1 Research Questions

1.1 Do advanced theory of mind tasks provide developmentally appropriate measures for able autistic 6-12 year olds by discriminating performance from matched typically developing controls in a way that standard theory of mind tasks do not?

1.1.1 Standard Tasks

Children performed at or near ceiling on the standard theory of mind measures used in this study. The Sally-Ann and Smarties false belief tasks and the Picture Sequencing task did not discriminate between the group of autistic children and typically developing children, matched for chronological age, verbal age and IQ. This contrasts with early studies that reported an autism specific deficit in performance on these theory of mind measures (Baron-Cohen et al., 1985; Baron-Cohen et al., 1986; Perner et al., 1987).

Baron-Cohen et al. (1985) had previously found that of 20 autistic children only 20% passed the Sally-Ann task, compared to 85% of typically developing children. The difference between these findings and current results may be accounted for by the difference in participants' characteristics. The autistic group in Baron-Cohen et al.’s
(1985) study had a lower verbal age (mean 5 years 5 months) than in this study (mean 8 years 3 months). The Baron-Cohen et al. (1985) autistic group also had a lower IQ (mean 82) than in this study (mean 100). The autistic group in the Baron-Cohen et al. (1985) study was also chronologically older (mean 11 years 11 months) than the group in this study (mean 9 years 4 months). This reflects a slower rate of language acquisition in Baron-Cohen et al.'s (1985) study than for the autistic group in this study, in addition to the lower functional ability.

In Baron-Cohen et al.'s (1986) Picture Sequence task, autistic children performed better than typically developing controls on mechanical stories (possibly because of the autistic group's higher non-verbal age), identically on behavioural stories, and worse on mentalising stories. Children with autism in the present study, however, performed similarly to typically developing controls on all three story-types. The difference between the studies might again be accounted for by the lower verbal age and IQ, but higher chronological age of participants from Baron-Cohen et al. (1986) than in the current study compared to.

The good performance of the autistic group in this study was as expected for an intellectually able group of 6-12 year olds with good language skills. They performed at an 80% pass rate for the Sally-Ann task and 95% pass rate for the Smarties task. There has been much interest in evidence for a relationship between verbal ability and performance on the standard theory of mind tasks (Happe, 1995; Dahlgren and
It may be that children’s better performance in the current study was associated with their higher verbal abilities. Although the results do not offer a causal explanation for this relationship, it seems likely that there is a bootstrapping in the development of the two skills (Charman and Shmueli-Goetz, 1997). Better verbal skills may enable children to share their own thoughts and discover that they may differ from other people’s, and children with an awareness of the difference between minds might be more motivated to develop communication skills to explore them.

1.1.2 Strange Stories

The Strange Stories task was effective in discriminating between the group of high-functioning autistic children and the typically developing children in the current study. The differences between groups were in the complexity of explanations given to account for the mentalising stories, and in the use of mental state language in their responses. There was also a significant difference in the number of incorrect mental states attributed to story characters. No significant differences were found in terms of the complexity of answers for the Stories that only involved physical or behavioural exchanges.

The Strange Stories were designed by Happé (1994) as an assessment of children’s ability to understand other people’s thoughts, feelings and desires. Children were asked questions about story characters’ motivation for saying things or behaving in certain ways. The tool was intended to provide an assessment of mentalising skills needed to
understand subtle interpersonal behaviour embedded in real life social situations. The aim was to identify difficulties associated with autism, while controlling for general cognitive delay, by comparing performance with that of typically developing children matched for chronological age, verbal ability and IQ.

Children with autism gave less complex answers than their typically developing counterparts for the mentalising stories, accounting for fewer of the implicit reasons for characters’ behaviour. However, they gave as much relevant information as typically developing children on the physical stories. A repeated measures analysis indicated that there was no difference in the general level of difficulty of the two types of Story. There was, however, a specific pattern of impairment by the autistic group on mentalising items.

This suggests that the difference in complexity scores for the mentalising items was not the result of difficulties in understanding the story or in articulating responses to questions generally. This is consistent with the hypothesis that children with autism have specific difficulty thinking and talking about mental states (Baron-Cohen, 1992). This difficulty for autistic children between 6 and 12 years with cognitive ability within the normal range is picked up on the Strange Stories task, but not by false belief or picture sequence tasks for individuals of this age and ability.

As a group, the children with autism used significantly fewer words that referred to story characters’ thoughts, feelings or desires in their responses to the Stories with mentalising
content than the group of typically developing children. This was consistent with the hypothesis that autistic children are impaired in their use of mental state language compared to matched controls (Tager-Flusberg, 1995). If the use of mental state language does represent the onset of an ability to think about minds (Shatz et al., 1983) it is not surprising that both the complexity and mentalising measures showed associated impairments compared to the typically developing group.

In Happé’s (1994a) original study, performance on the Strange Stories clearly distinguished between autistic children unable to pass false belief tasks and typically developing children. The more able group of autistic participants in Happé (1994a), who had a mean VIQ 95, were more comparable to the autistic sample in this study than the ‘no theory of mind’ and ‘first order theory of mind’ sub-groups from the Happé (1994a) study. The Happé (1994a) second-order theory of mind group’s performance was indistinguishable from typically developing children but showed differences from typically developing adults. Conclusions from Happé’s (1994a) comparisons of the autistic group with these two typically developing groups were somewhat unclear because groups were not matched. In this study, direct comparisons between the autistic group and typically developing children were matched on chronological age, verbal ability and IQ.

1.1.3 Cartoons

There was no significant difference in performance by the autistic and the typically developing groups on either the mentalising or physical cartoons, or between the number
of incorrect answers given by the two groups. This was an unexpected result in view of
the difference between performance on the mentalising and physical stories. The Cartoons
had been designed to provide a measure of mentalising ability with reduced language
loading on the task stimuli. This was partly in response to criticisms of the Strange
Stories that performance was confounded by the language ability required to read and/or
understand the text. However, both tasks required children to give a verbal explanation of
their understanding of the stimuli, be they stories or pictures.

The similarity in performance between the two groups on the Cartoons cannot be
explained by performance at ceiling. In fact, the autistic group scored 53% of possible
marks for the mentalising items and 55% for the physical items, compared to 58% of
mentalising and 60% of the maximum score for physical items by the typically
developing group. This indicates that neither group explained why the Cartoons were
funny as fully as they might have done. It is possible that the question, "What is funny?"
was not sufficient to access children's concepts of other people's minds when a simplistic
answer appeared to suffice. For example, children looking at the Cartoon with a bull
charging towards a toreador (Cartoon M3) explained that what was funny was the
presence of an iron block, without referring to the effect it was about to have.

Alternatively, the Cartoons might not have clearly distinguished between mentalising and
physical skills as clearly as the Strange Stories did. For example, the picture of mice
setting up chairs to watch TV (Cartoon P6) included a rather surprised-looking man
coming into the room. The recognition of his mental state might have been important in understanding that it was a funny thing for the mice to be doing, even though this item was categorised as a physical cartoon. Overlap between mentalising cartoons that could be answered in physical terms and physical items that required an understanding of mental states might have weakened the differential strength of the tool.

1.1.4 Eyes task

There was a significant difference between the autistic and typically developing groups’ ability to match target words to the facial expressions in the Eyes task. The results were consistent with Baron-Cohen et al.’s (1997) finding that there was a significant difference in the performance of autistic and typically developing participants. However, in the original study, only half the autism/Asperger syndrome group (with IQ within the normal range) performed above chance, while in the present study all participants performed above the level of chance. Differences between the two studies might have partly been because of changes from the original adult task to the child task used in this study (Baron-Cohen, unpublished). The Baron-Cohen et al. (1997) task required adults to make a forced choice between two opposite words, while in the children’s version a choice was made from four unrelated target words.

1.1.4 Advanced Measures

As a whole, the tasks were able to predict group membership at an 85% accuracy level. Comparing the ability of the various tasks in discriminating between the two groups, the
most effective predictor of group membership was the complexity score on the Strange
Stories task. The score for the use of mental state language of varying levels and
performance on the Eyes task were also effective discriminators between the two groups.
Differences in performance on these measures was not dependent on chronological age,
verbal age or IQ. The Cartoon task was not a useful tool for distinguishing the two
groups.

1.2 Are the difficulties in theory of mind task performance associated with autism
more usefully described as a continuum of impairments, rather than an absence of
mentalising ability?

1.2.1 Mentalising Abilities in Autism

There was a significant difference in the use of mental state terms on the Strange Stories
task between the two groups. However, this should not be interpreted as a complete
absence of mentalising ability in the autistic group. In fact, the autistic group used some
mental state language in 6 of the 8 explanations for the mentalising stories. The variation
of performance within the autistic group reveals that although mentalising ability
distinguishes the two groups as a whole, there are individuals who are able to produce
answers with an awareness of mental states similar to the typically developing group
(Charman, 1998).
Initially, it was proposed that there was an autism specific deficit in the ability to pass false belief tasks (Baron-Cohen et al. 1985). This was explained by the absence of a decoupling mechanism necessary for the formation of meta-representations of other people’s thoughts (Leslie, 1988). Since then there has been an attempt to account for the consistent proportion of autistic participants who seem able to pass first and second order false belief tasks.

Frith et al. (1991) suggested that autistic individuals with verbal and intellectual ability within the normal range might have developed cognitive strategies to compensate for their lack of theory of mind abilities. This could enable them to ‘hack-out’ answers to theory of mind tasks but might not be comprehensive enough to facilitate social interaction in real life settings. This hypothesis was supported in the current study by the finding that the autistic group made more incorrect mental state attributions than the typically developing group. For example, one autistic participant suggested that the girl given an encyclopaedia instead of a rabbit for Christmas was happy to receive it because it might contain pictures of animals, rather than demonstrate an understanding that she was lying to protect her parents’ feelings.

It might be that although some members of the autistic group knew they should account for the Stories in terms of the character’s motives, they needed to think up an answer in a problem solving way rather than simply report their knowledge of a character’s predicament. There may be some scenarios which are relatively common to children for
which autistic children may have learnt to apply a rule, such as learning to say only nice things to people so as not to hurt their feelings. It may be that in some situations children might apply complete stock phrases rather than use novel word combinations in new situations.

Some attempt has been made to identify what evidence would distinguish 'hacked-out' answers from 'real mentalising.' For example it might be that 'hacked out' responses are associated with a slower response speed or higher IQ scores. Children with IQ within the normal range might be more able to use alternative problem solving strategies to compensate for impaired mentalising abilities than those with additional cognitive delay (Yirmiya et al., 1992).

In this study, mentalising ability in the autistic group was not correlated with IQ, but it might be that for a group of autistic children who have a wider range of IQ this variable would be a discriminating factor between high and low performance. Bowler (1997) compared the time taken by autistic and typically developing children matched for IQ to complete second order mentalising and non-mentalising tasks. Although the autistic group were slower than matched controls, there was no difference between time taken on mentalising and non-mentalising items for either group, which did not support the hypothesis that different strategies were used for processing mentalising and non-mentalising information.
An additional explanation might be that autistic individuals have some mentalising abilities, even if they are impaired to varying degrees. There has been a move in the field from non-existant mentalising ability to a continuum model (Tager-Flusberg, 1996). The pass/fail design of standard false belief tasks supported a presence/absence interpretation of theory of mind data, while advanced tasks used in this study measure the performance on a scale. This enables the use of parametric analyses, which were considered appropriate by Buitelaar et al. (1999) in the light of participants’ non-bimodal pattern of performance.

Frith et al. (1994) suggested that there might be sub-groups matching each of the two hypotheses. There seems to be some support for both explanations in the current sample of autistic children. Some children’s responses scored points without demonstrating a full understanding of characters’ mental states. For example, asked why Mrs Smith said she would have to drown her kittens (Story M4), one autistic child explained it was “because she hated them, perhaps.” In contrast, other answers clearly demonstrated an understanding of the motivations that lay behind story characters’ actions. An autistic child explaining the story of the greedy schoolboy (Story M3) said, “He thinks he’ll be cunning and say that he won’t have dinner,” while another child talking about the captured soldier (Story M2) said, “He was clever and didn’t panic. They thought he would lie and would go to the sea.”
Interestingly there was also variation in the responses made by the typically developing children. Some participants responded with mentalising statements that did not demonstrate an understanding of the implicit message of the story. For example, one typically developing child explained the soldier's double bluff (Story M2) by saying, "He wanted them to know where the tanks are." Another said the reason the burglar (Story M8) gave himself up when stopped by the policeman was "because he wanted the glove." These stories were originally designed for typically developing 8-9 year olds and it might be that the wider age range in this study means some children are too young to perform as expected. This explanation is consistent with the strong correlation between story performance and chronological and verbal age.

1.2.2 Levels of Mentalising Complexity

The results from this study indicate that autistic children with language ability and IQ within the normal range differ from typically developing controls in their ability to understand second order mentalising stories, but perform similarly on first order items. Both groups performed similarly across the physical story items, finding it most difficult to explain why the tanks beat the airforce (Story P1), and easiest to explain why the tree fell in the storm (Story P3). This reflects the variation of the cognitive challenge posed by the different items, which was similar for both groups.

Both groups were competent in the appropriate use of basic social conventions, such as being polite to relatives (Stories M5 and M6). The items which differentiated the two
groups were more complex mental state attributions, particularly understanding that one character does/does not have a correct understanding of another character's thoughts. For example, to answer the story about the captured soldier (Story M2), children needed to understand what the enemy army thought he would think. It may be that autistic individuals are able to develop basic skills but do not go on to acquire higher order capabilities, unlike their typically developing counterparts.

There was also a difference in the use of higher order mental state language between the two groups, while there was no difference in the use of first order language. For example, second order statements such as, "He was worried that she would be upset," were used more frequently by typically developing children, while autistic children were more likely to use first order explanations such as, "She might be upset." Buitelaar et al. (1999) found second order tasks were the most appropriate tools for autistic children with intact intellectual abilities who performed at ceiling on first order tasks.

Of the Cartoons, there was one item (Cartoon M4) which autistic and typically developing children answered differently, but it is unclear why this cartoon differed from the others. Similarly it is unclear whether there was a pattern to the items on the Eyes task which distinguished the groups. For both tasks, items which pulled the groups apart were only significant at p < .05, while a probability of p < .01 would be more appropriate considering the repeated analyses on a relatively small data set.
Items on the Eyes were not classified according to the type of target words, leaving it unclear whether there was a difference in performance for types of emotions. Poor performance might be because of problems understanding the expression, other factors in the photographs, the target words or the distracters. One of the items was associated with higher scores for the autistic group than the typically developing group, although it seems unclear what explanation might account for this.

There is a suggestion in the literature that autistic children are able to match simple emotion words, such as “happy” and “sad,” while they have difficulty with emotion words dependent on a theory of mind, such as “surprised” (Capps et al., 1992). It is unclear, however, whether matching target words to photographs in the Eyes task is similar to matching pictures of emotions, or if the task was tapping another underlying skill.

1.2.3 Individual Differences

Although there was a significant difference between group performance on the mentalising stories, the spread of performance within the groups indicates an overlap between the performance of some autistic and some typically developing children. In the autistic group, 5 children scored 12 or more out of a possible 16 for the complexity scores. In the typically developing group one member scored only 4 out of 16, but the majority scored at least 14 out of 16.
There was a wide range in performance across mentalising and physical cartoons for both groups. On the Eyes task, there was some clustering in the autistic group, with all but 4 participants scoring 10-15 correct out of a possible 27. The pattern largely reflected the distribution of verbal age in this group. The spread in the typically developing group was from the same lower level but with a higher upper limit, with scores ranging more evenly between 10 and 23 out of 30. These differences were not accounted for by verbal age or IQ.

The pattern of a decreasing proportion of children with autism with increasing mentalising ability (Sparrevohn and Howie, 1995) was found in this research, and was consistent with Happé's (1991) study.

1.3 Are there characteristics that are associated with mentalising ability in autism?

In the present study there was no relationship between either story complexity or mentalising language, with chronological age, verbal age or IQ for the autistic group. This contrasts with findings from previous studies, which have found strong correlations between verbal ability, in particular, and performance on theory of mind tasks (Eisenmajer & Prior, 1991; Happé, 1995). Charman and Baron-Cohen (1992) argued that high verbal mental age was required but not sufficient to pass false belief tasks. The majority of studies which found a strong relationship between verbal age and theory of
mind ability were for groups of autistic children who were less able than in the current study, and tested with standard false belief tasks (Happe, 1995). It might be that beyond a certain level of ability verbal age is a relatively weak predictor of advanced task performance.

In contrast, performance on the Eyes task was strongly correlated with verbal age in the autistic group. This might be explained by the relative complexity of target and distractor words in the study that could be difficult even for children who have average verbal ability. It might be that performance on this task for autistic children is more a measure of the ability to understand and use words such as “ashamed,” “excited” and “daydreaming” than of mentalising ability. An alternative explanation might be that the Eyes task measured an unknown skill that strongly correlated with verbal age for autistic children, but not with performance on the Strange Stories. It is an assumption that the Strange Stories offer a more useful measure of mentalising ability, and that this would imply that the non-related performance on the Eyes tasks measures a non-mentalising skill. It could be that the Eyes measures some mentalising ability which is not picked up by the Strange Stories tasks.

Verbal age correlated with the Strange Stories complexity scores, and the Cartoons mentalising score for the typically developing group. This is consistent with previous findings that measures of verbal ability are associated with theory of mind ability in typically developing children (Happe, 1995). It suggests that the relationship is not simply
a reflection of word learning increasing with theory of mind, but a more general reflection of language skills, since verbal ability was measured by four CELF-R sub-tests in the current study rather than the BPVS.

Previous studies have found a relationship between IQ and autistic children's ability on false belief tasks. This was consistent with the view that more able children can pass tasks by using a problem solving strategy as a substitute to using impaired mentalising abilities (Yirmiya et al., 1992). Although there was a range of performance on theory of mind tasks within the group, performance was not associated with this variable, except for the physical stories complexity score. There was no relationship between IQ and task performance for the typically developing group. It might be that other factors play a role in the development of theory of mind development, such as opportunities to learn about and practice social behaviour (Dunn, 1996).

1.4 Is there evidence to support a distinction between ‘cognitive’ and ‘affective’ aspects of mentalising abilities?

1.4.1 Types of Mental State Language

This study found no significant difference in the use of desire, emotion or cognition words between the two groups, although there was an overall difference in the use of mental state language in the Strange Stories task. There is an assumption that the use of
mental state language signifies a developing understanding of other people's minds, although it is unclear whether children do fully understand the meaning of the words from the time they start to use them (Shatz et al., 1983).

The results of this study do not support some previous findings that desire words are used similarly by autistic and typically developing children, in contrast to impaired use of some emotion words, and most cognitive language (Tager-Flusberg, 1989; Capps et al., 1992). This argument would predict that children with autism followed a delayed developmental pattern of mental state language acquisition (Ozonoff et al., 1991). In typical development words such as “want” or “like” emerge first, followed by simple emotions such as sad and happy. More complex emotions, such as “surprise” and “shame,” which are dependent on an understanding of associated thoughts, might not be acquired until the third year, with the development of cognitive words such as “think,” “know” and “believe” (Wellman, 1990).

This study found an overall impairment in the use of mental state language the autistic group, but no pattern according to the type of language. This was similar to a pattern of overall impairment in emotion tasks, but no difference between simple and complex emotions, found by Buitelaar (1999). In both studies, the autistic group had a high level of verbal ability which suggests participants might have reached the necessary stage of development to acquire these skills, even if they were used less well than by the typically developing group.
1.4.2 Task Loading of Cognitive and Affective Skills

Performance on the Eyes task was significantly poorer for the autistic group than the typically developing group, even though children with autism demonstrated their ability to match target words to photographs well above chance levels. It is difficult to compare performance on this task with previous findings because it remains unclear what ability the task measures.

Scores on the Eyes task did not correlate with performance on the Strange Stories or Cartoon tasks. It might be that the Eyes task measures a different skill to that picked up by the Stories and the Cartoons, which also discriminates between the two groups, such as some form of emotion reading skill. It is unclear whether the Eyes test is a measure of the ability to recognise emotions or feelings in other people's faces, or an ability to attribute thoughts to other people's minds. The task differs from emotion reading/recognition tasks which, for example, might ask participants to match a photograph of an expression to related pictures of the same emotion from a set of distracters (Ekman et al., 1972).

Although Baron-Cohen et al. (1997) claimed that performance on the Eyes 'mirrored the pattern of performance on the Happé Strange Stories' (Baron-Cohen et al., 1997, p.820), there seems to be insufficient data to conclude that the Eyes task is a measure of theory of mind. There was no relationship between performance on the Eyes and on the Happé
Strange Stories in the current study, although it is not possible to conclude which, or even whether either, is a reliable measure of mentalising ability. An explanation might be that the language used for the children’s Eyes task pose more of a challenge to verbal ability than the original stimuli did for adults (Baron-Cohen et al., 1997). Although the task was designed to have minimal language demands, by using photographs and simple verbal instructions, the four target words were relatively complex.

Children seem able to match emotion words to faces before they become fully aware of the thoughts people experiencing those emotions might have, for example knowing tears mean a sad face before they understand the reasons why people might feel sad. Even so, there seems to be an association between affective and cognitive development, more integrated in typical development than in autism (Buitelaar and van der Wees, 1997). Although performance on the Eyes and Strange Stories tasks were not related in this study, no conclusions can be drawn as it remains unclear what skill the Eyes task measured.
2 Methodological Issues and Clinical Implications

2.1 Measures of Participant Characteristics

Autistic children assessed at the Children's Centre are routinely tested on the K-ABC and CELF-R. In the light of ethical issues of administering a long and unnecessary test by using the WISC after a K-ABC, and the possibility of a practice effect for the CELF-R, language and IQ tests were only administered for autistic children who had not been assessed in the preceding 12 months. Results from the initial assessment by staff at the Centre were used for those who had been more recently assessed. Despite this, children completed a long assessment of up to 3 hours, which might result in under-estimations of performance on language and IQ scales which were administered after the theory of mind matures, because of fatigue.

Although the literature supports a correlation between the K-ABC and WISC, only a full-scale score could be calculated as the K-ABC did not offer separate verbal and performance IQ. Care was taken to match children for verbal ability on the CELF-R, but there was no data to match participants on non-verbal abilities. Some studies have found a relationship between non-verbal mental age (e.g. Tager-Flusberg and Sullivan, 1994) while others have not (Charman & Baron-Cohen, 1992).

There was an unequal proportion of males and females in the two groups. It was not surprising that the autism group was characterised by more males than females, but there
was a bias in the opposite direction with the majority of the typically developing group being female. Baron-Cohen et al. (1997) suggested that females were better at mentalising than males, but this was based on a relatively small data set. Charman et al.'s (1999) analyses of a much larger sample indicated that females' better performance on false belief tasks was limited to younger children. Although it could be argued that the difference in performance between the two groups might be influenced by the good performance of the typically developing females compared to the poorer performance of autistic males, gender comparisons within each group in this study did not identify differences in performance. However, small and unequal numbers within each group limited the strength of comparisons, so that future work would benefit from matching for gender as well as chronological age, verbal age and IQ.

2.2  Mentalising Measures

The Strange Story responses were marked on two rating scales, a complexity rating of 0-3 and a mentalising score. Great care was taken in ensuring the measure’s reliability by double marking and carrying out intra-class correlations, which indicated a high level of reliability. In marking the responses, however, it became apparent that the two measures were confounded because most mentalising items required an understanding of characters’ mental states in order to achieve a maximum score on the complexity measure. In contrast, the responses on the Cartoons were marked with a single score in which the complexity maximum score reflected an understanding of characters’ mental states.
The physical stories were intended to provide a control to identify any performance difficulties that were not associated with mentalising abilities. It might be that the reason why there was no difference between the autistic and typically developing groups' responses for these stories was that they required simpler explanation. The scoring criteria on the physical stories was less specific than the mentalising criteria. For example, asked why the clothes were wetter than they had been (Story P3), a correct answer was simply that it had been raining. In contrast, a full answer for the double bluff (Story M2) required a response explaining that the soldier knew that the enemy would think that he would lie and therefore they would plan to go to the opposite location to the one he said.

It might be that a more useful measure for autistic children with high verbal and intellectual ability would omit items on which both groups scored highly. These would be Stories that only required low-level mentalising skills, or the use of alternative cognitive strategies. More items of the type that differentiated the two groups, which depended on the use of higher level mentalising skills might be added and tested for reliability and validity. This is consistent with the recommendation of Buitelaar et al. (1999) that second order tasks are more useful in differential diagnoses, as most children of 8 years and above with verbal and intellectual skills intact perform at ceiling on first order tasks.

The children's Eyes task is an unpublished adaptation of the adult version (Baron-Cohen et al., 1997) which differs in the use of three unrelated distracters and a correct target
word, rather than a forced choice between two opposites. The latter was criticised for the ease of selecting a general idea of positive or negative emotion. A difficulty in the updated system is that errors might be associated with confusion about the target and distracter word meanings rather than reflecting a difficulty in recognising the emotion expressed by the Eyes.

2.3 Generalisability of Findings

The 'advanced' theory of mind tasks used in this study were unpublished and therefore the comparability of this study to previous research is limited. Previous versions of the Strange Stories and Eyes tasks have been used although the scoring criteria for both have been amended by their creators. Care was taken to use the criteria developed by and in collaboration with the authors to enable comparisons to be made with any future studies using these tasks. Any further work needs to address the test-retest reliability of these unpublished measures in the light of mixed findings from test-retest reliability on standard theory of mind tasks (Mayes et al., 1996).

Previous studies have distinguished between participants with autism and Asperger syndrome, although this study used a single group composed of children with diagnoses under the umbrella of the spectrum of autistic disorders. This decision was based on studies such as Dahlgren and Trillingsgaard (1996) and Prior et al. (1998). They found that sub-groups within the research match diagnoses of autism, Aspergers syndrome and PDD-NOS in terms of the range of severity of characteristics, but not in any clinical
A criticism of the approach taken in this study could be that performance of the autistic group might have been confounded by the inclusion of children with subtly different disorders. Ziatas et al. (1998) found a significant difference in theory of mind performance between their autism and Asperger syndrome groups. However, this might have been because of the significantly lower verbal mental ages associated with differences in children's severity of symptoms, which distinguished between passers and failers, rather than a qualitative difference in their diagnoses. Similarly, there was no evidence that the high pass rate of Asperger syndrome adolescents in Bowler's (1992) study reflected any difference from the performance of autistic children in previous studies, other than their significantly higher verbal mental age. It seems that verbal age close to chronological age, which has been used a diagnostic criteria for Asperger syndrome, is associated with mentalising abilities.

The autistic participants in this study were recruited from a tertiary referral centre, known to specialise in the diagnosis of pervasive developmental disorders. It might be that a better knowledge of autism has resulted in an increase in diagnosis by paediatricians at the second tier, which has consequences for services such as the assessment centre from which children were recruited. It may be that there was a selection bias towards an atypical presentation of more able autistic children who pose difficulties for diagnosis at second tier services which might account for the high scores across the tasks found in the
This might limit the generalisability of findings in this study to other autistic children, who generally have much poorer verbal and intellectual skills. The majority of researchers view an autistic group with IQ and verbal age within the normal range advantageous in providing a 'pure' presentation (e.g. Buitelaar and van der Wees, 1997). However, Bailey et al. (1996) emphasised that the high co-morbidity between autism and learning disability (75% IQ < 70) suggests that there is a relationship between the two which should not be ignored.

Alternatively, there may be particular validity in investigating the performance of autistic children who require detailed assessments for differential diagnoses. Children who present with classical autistic features across the triad of impairments do not require the development of more sensitive measures. Until now, theory of mind tasks have mainly remained within the research arena as they have not been sensitive enough to offer more clinically useful information than existing diagnostic tools. It might be useful for the Strange Stories task to be amended in order to become a more useful diagnostic tool by excluding first order scenarios, replacing them with additional second order story items, to offer a more sensitive measure of high level mentalising skills.

In order to demonstrate clinical utility, future research should examine performance on an updated Strange Stories between clinical populations, particularly those who tend to
overlap on differential diagnoses such as SPD and ADHD. This would be of particular interest in the light of recent findings that children with ADHD were not distinguished from matched children with autism on 'standard' first and second order theory of mind measures (Buitelaar et al., 1999). However this finding must be viewed tentatively in the light of Charman et al.'s (1999) contrary finding that children with ADHD performed equivalently to typically developing children on the Happé (1994a) Strange Stories.

It may be that tools such as the Strange Stories could be an aid for social skills training as well as differential diagnosis. If language difficulties are associated with impairments in the ability to draw on preserved skills, current speech and language therapeutic approaches might usefully be supplemented by interventions around the use of language in real life social settings. The tools could offer practice situations to plan strategies to deal with confusing situations and to assess children's progress.

2.4 Impairments in mentalising competence or performance?

Children's explanations for the Strange Stories and the Cartoons were not prompted after the initial question and should be seen as explanations which they considered appropriate to answer the question, rather than as an indication of the extent of their understanding. Rather than demonstrating an inability to understand why story characters behaved as they did, the difference between the performance of the two groups might be more a reflection of the difference between the two groups' perception of what the experimenter needed to know. Bowler (1992) commented that the nature of theory of mind tasks
influence participants’ responses, and that performance should not be an indication of competence. Proponents of the theory of mind have emphasised the nature of tasks as ‘remote probes’ for mentalising skills rather than accurate measures of underlying ability (Frith, 1992).

It might be useful to explore children’s understanding in a more flexible way. Perhaps implementing a qualitative study using structured interviews could provide new information about autistic children’s ability to think about other people’s minds. Previous studies suggest that in more structured settings (particularly with familiar people) autistic children can demonstrate greater competence in conversing than they do in unstructured narratives (Tager-Flusberg and Anderson, 1991). In the current study, for example, it was unclear whether the autistic child who explained that the cat owner said she would drown her cats (Story M4) "because she could not keep them," thought that she would really do it. The child might have been aware that the woman could not keep them but would not drown them even though he did not give a full, unprompted answer.

Children in this study who performed well on the mentalising tasks fulfilled criteria for autism, experiencing difficulties across the triad. A weakness of the theory of mind hypothesis has been in reconciling pervasive social, communicative and imaginative impairments with a proportion of autistic individuals’ ability to pass standard mentalising measures. The difference in performance on ‘advanced’ tasks between autistic and typically developing children in this study partly supports Baron-Cohen et al.’s (1997)
explanation that tasks are limited in simulating real life scenarios that are often more complex and subtle, amidst greater distractions and pressures.

Bowler (1992) and Buitelaar et al. (1999) suggested that the mentalising ability of verbally and intellectually able individuals with autism, despite their profound difficulty in real life settings, might result from a difficulty in using available resources effectively. Additional cognitive hypotheses have been proposed to account for the difficulty in using available skills. Frith (1989) suggested that weak ‘central coherence,’ which normally acts to draw information together in a drive for meaningfulness, results in the fragmentation of information for individuals with autism. Irrespective of performance on theory of mind tasks, children with autism were found to be less likely to use the context of a story to decide on the pronunciation of homographs than typically developing children (Happe, 1991). It might be that autistic individuals have difficulty in using a theory of mind if it does not draw on previous experiences and the context of new information (Happe, 1994).

There is some overlap between the proposition that difficulties responding to the relevant context, and problems adapting to change associated with executive function difficulties (Frith and Happe, 1994). Both accounts demonstrate that difficulties associated with autism are not limited to the mentalising domain (Ozonoff et al., 1991). Russell et al. (1991) used a Windows task to demonstrate that children with autism, like typically developing children under 3 years old, persisted in pointing to the real location of a
chocolate rather than demonstrate an ability to ‘disengage from the object’ in order to keep the treat for themselves.

However, it seems that these theories provide additional explanations rather than an alternative to the impairment in mentalising (Frith and Happé, 1994). Autistic children’s difficulties in this study were specific to the mentalising stories. Leslie and Thiass (1992) demonstrated that autistic children who saw an instant photograph taken of a doll sitting on a bed, correctly predicted where the doll would be sitting in the picture after it had been moved onto a rug. They differed from typically developing children under 3 years who reported the current position of the doll rather than the original situation recorded by the photograph (Zaitchik, 1990). It might be that the characteristics associated with autism result from an interaction between the varying degrees of an underlying organic impairment in the ability to think about minds, with an additional impairment in the ability or motivation to use preserved skills (Bailey et al., 1996).

The question remains why there would be a motivational deficit in use of mentalising information. Hobson (1993) has suggested that the answer may lie with early disturbances in the experience of interpersonal contact. If children with autism did not experience the innate positive arousal associated with the infant-carer connectedness they might differ from typically developing infants who seem motivated to engage in activities with other people that re-create that experience. Typically developing children develop protodeclarative gestures which appear to have the purpose of sharing an interest
with another person, as distinct from protoimperative gestures which function as a request for something tangible. It has been suggested that differences in shared attention and pretend play, which are dependent on the beginning of an awareness of the difference between people’s viewpoints, may identify early indicators of autism (Happé, 1994).

Concluding Remarks

The autistic and typically developing groups were carefully matched according to age, verbal ability and IQ, which was reflected in their similar performance across the non-mentalising control items. There was, however, a specific difference between the two groups in the ability to explain characters' behaviour in second order mentalising Strange Stories, and in the use of second order mentalising language in their responses. This is consistent with previous research that suggests autistic children with IQ and verbal abilities within the normal range have a specific impairment in demonstrating higher order mentalising abilities possessed by matched non-autistic children.

In an attempt to develop theory of mind tasks that do not pose such a heavy verbal demand as the Strange Stories, two further advanced tasks had been developed, although not yet published. This study found, however, that the Cartoons task did not provide a useful tool in differentiating between autistic and typically developing participants. Although the Eyes task did differentiate between the two groups, it remains unclear what skills it measures. Unlike the adult Eyes task, this research did not find that performance
was correlated with the Strange Stories, and therefore it cannot be confidently be described as a theory of mind measure, assuming the Strange Stories offers some measure of mentalising ability.

This study supports the view that mentalising is not an all or none affair but would most usefully be described as a continuum of ability. It seems that some autistic children with good verbal skills and IQ within the normal range can demonstrate some ability to explain why people behave in certain ways, and use mental state language to do so. It remains unclear whether this is the consequence of intact mentalising competence or the implementation of alternative strategies. What seems important, however, is that even in a group of high functioning children who performed similarly to typically developing controls on 'standard' theory of mind tasks, there was a significant difference in performance on two of the three 'advanced' theory of mind measures used here. This suggests that the development of measures incorporating second order theory of mind skill could be useful in differential diagnoses and social skills taining of able autistic individuals.
References


Charman, T. (1998a) Specifying the nature and course of the joint attention impairment in autism in the pre-school years. Implications for diagnosis and intervention. Autism, 2, 61-79.


and neuropsychological characterization of Asperger syndrome: convergence with non-verbal learning disabilities syndrome. Journal of Child Psychology and Psychiatry. 36, 1127-1140


Mahoney, W.J., Szatmari, P., MacLean, J.E., Bryson, S.E., Bartolucci, G., Walter, S.D.


Appendices

1. Ethical Approval

The University College London Hospitals

The Joint UCL/UCLH Committees on the Ethics of Human Research

Committee A Chairman: Dr F D Thompson

Dr T Charman
Senior Lecturer in Clinical Psychology
Institute of Child Health
UCLMS
30 Guilford Street
London WC1N 1EH

27 June 1998

Dear Tony

Study No: 98/0076
Title: Assessing the role of language ability in theory of mind performance in children with autism and language disorder

Many thanks for your letter dated 17th June and the revised information sheet. Dr Thompson has no further objections to this study going ahead.

Please note that it is important that you notify the Committee of any adverse events or changes (name of investigator etc) relating to this project. You should also notify the Committee on completion of the project, or indeed if the project is abandoned. Please remember to quote the above number in any correspondence.

Yours sincerely

Iwona Novicka
Administrator, UCL/UCLH Ethics Review Committees
CONFIDENTIAL

CLINICAL GROUP INFORMATION SHEET - (a) AUTISM
Theory of Mind Assessment Battery
May 1998

You and your child are being invited to participate in a research project which aims to develop an extension of the clinical assessments which have been used for several years at Harper House for children with developmental difficulties. It is hoped that it will be useful in giving clinicians a better picture of individual children's abilities and difficulties, and how best to support them.

You are being asked to participate because your child has been referred to Harper House Children's Service because of his/her developmental difficulties, and is between 6 and 12 years of age. The research is being carried out by Dr. Tony Charman, Lecturer in Psychology, at the Institute of Child Health, and Ella Brent, Clinical Psychology Trainee at University College London.

The tests involve asking your child about their understanding of a set of stories where a character makes an ambiguous statement, about a set of cartoons depicting social and slapstick jokes, a task involving the identification of emotions from photographs of facial expressions, and providing a word to complete a sentence where only one of 2 pronunciations of words would apply. The tasks are presented as games, and it is emphasised that there are no right and wrong answers. Previous work with similar materials have shown that most children find the session fun.

Please read this form carefully and retain it for your information.

If you and your child decide to participate in this study, the following will apply:

1. One of us would like to visit you at Harper House, to complete a series of language and social skills tasks with your child. We would also like you to answer some questions about your child's difficulties and abilities. The session will last about an hour in total.

2. All the information that is collected from you and your child will remain confidential to the research team. No individuals details of any child's performance will be released.

3. You and your child do not have to participate in this study if you do not want to. If you decide to take part you may withdraw at any time without having to give a reason. Your decision whether to take part or not will not affect your child's care or management in any way.

All proposals for research using human participants are reviewed by an ethics committee before they can proceed. This proposal was reviewed by the Ethics Committee at the Joint UCL/UCLH Committees on the Ethics of Human Research. If you have any questions please do not hesitate to contact Ella Brent on 0171 or Tony Charman on 0171

Ella Brent
Clinical Psychologist in Training
University College London

Dr. Tony Charman
Lecturer in Psychology, Institute of Child Health
Honorary Chartered Clinical Psychologist
Harper House Children's Service
March 1999

Dear

I am writing to you to ask for your help in a research project organised by University College London. The aim of the research is to help us in the assessment of a rare developmental disorder called autism. It is a disorder that profoundly affects children’s ability to socialise, communicate and play. I am looking for participants to form a comparison group of children who are non-autistic but have some language difficulties.

There is a theory that the reason for these problems is that children with autism find it hard to understand that people have thoughts, and often appear to treat people more as if they were objects. Several tasks have been designed in the form of enjoyable games to see how children can ‘put themselves in other people’s shoes’ and imagine how they might think in different situations.

The reason for our research project is to find out if some children with autism might actually find some of these games hard just because they do not understand the language used in the tasks, rather than just having difficulty with understanding how people behave in certain situations. Therefore we are comparing the performance of autistic children on tasks which use a lot of language with tasks that do not use much language at all.

In addition to the assessment of children with autism we need to see how children with language difficulties who are not autistic would do. Therefore I am writing to you to ask permission for me to assess your child as a comparison of how we would expect a non-autistic child of this age to do on the tasks.

The tasks are in the form of games which we have found children enjoy playing. They will be asked to predict how a story character is likely to behave in different situations, to describe the emotions shown in photographs of peoples faces, to suggest why story characters might say the things they do, and say what is funny about some cartoons.

I would like to assess your child at school at a time considered convenient by his/her teacher. I enclose a permission slip which I would very much appreciate you completing as soon as possible and returning to school. The assessment will be a one off which will take about 2 hours to complete. A copy of the overall findings of the study will be available at the end of the Summer. If you would like to discuss this further please do not hesitate to call me on 0171 435 7111.
All the proposals for research using human participants are reviewed by an ethics committee before they can proceed. This proposal has been reviewed by the Camden and Islington Community Health Trust Ethics Committee. All the information that will be collected is anonymous and confidential to the research team, and no details of any individual who takes part in this study will be released.

Your child does not have to take part in the study if you do not want him/her to. If you decide to take part you may withdraw at any time without having to give a reason. Please keep this information sheet for your further reference.

Thank-you very much in advance, for you co-operation which is very much appreciated.

Yours sincerely

Ella Brent
Clinical Psychologist in Training

Dr. Tony Charman
Lecturer in Psychology, Institute of Child Health
CONFIDENTIAL

CONSENT FORM FOR CLINICAL GROUPS

Theory of Mind Assessment Battery

This study will be carried out by:
Dr. Tony Charman
Lecturer in Psychology
Institute of Child Health

Ella Brent
Clinical Psychology Trainee
University College London

To be completed by parent

1. Have you read the information sheet about this study? YES/NO
2. Have you had the opportunity to ask questions and discuss this study? YES/NO
3. Have you had enough information about this study? YES/NO
4. Have you received enough information about this study? YES/NO
5. Who have you spoken to about this study? __________________________________
6. Do you understand that you and your child are free to withdraw from this study
   • At any time
   • Without giving a reason for withdrawing
   • Without this affecting his/her care, management or future education? YES/NO
7. Do you agree for you and your child to participate in this study? YES/NO
8. Do you give permission for your child's teacher/worker to be contacted? YES/NO

Signed _______________________________ Date _____________________

Child’s Name (block letters) ____________________________________________

Your Name (block letters) ______________________________________________

Researcher __________________________________________________________________

Head of School / Centre _________________________________________________

School / Centre __________________________________________________________________
5. Scoring Criteria for the Strange Stories

Examples for each Score. M= Mentalising Stories  P= Physical Story.

P1 Armies
2. The Green army is stronger in the sky, but the fog meant they lost their advantage, and therefore the Blue army won.
1. The Blue army is stronger on land.
0. Blue army is stronger in tanks. OR Tanks can shoot aeroplanes more easily.

M1 Brothers
2. Jim knows that Simon always lies and so when Simon told him where the bat was, he knew that it would really be in the other place.
1. Simon is a liar.
0. The bat was under the bed.

P2 Burglar
2. The burglar trod on an animal which ran through the detector beam and set off the alarm.
1. The animal set off the alarm
0. The burglar accidently set it off.

M2 Armies
2. The soldier knew that the enemy will think that he will lie and therefore they will look in the opposite place to the place he says. Therefore he tells them where his army really is, in the mountains so that the enemy will look in the sea.
1. The soldier tells them to look in the mountains so they will look in the sea. To trick the Red army.
0. He forgot where they were. OR He wanted the army to be caught.

M3 Sausages
2. He thinks that they will feel sorry for him if he says he wont have tea at home, and so they are likely to give him more sausages.
1. He wants to have more dinner. OR He is greedy.
0. Because he is hungry.

P3 Washing
2. It started raining once he had gone out.
1. It was raining.
0. It was not really.

M4 Kittens
2. She is pretending that she would drown them, even though she wont, in order to persuade the girl to buy one of her kittens.
1. To make her buy a kitten. OR Because nobody wanted to buy one.
0. Because she could not keep them.

M5 Hat
2. He does not want to hurt is aunt’s feelings.
1. She might not visit him again if he is rude.
0. It was horrible.
He did like it really.

**P4 Broken Leg**
2. To see if she has broken any bones.
1. To see if it was hurt.
0. Because the Dr. wanted it X-rayed.

**M6 Rabbit**
2. She does not want to seem ungrateful for her present / hurt her parents’ feelings.
1. She does not want her parents to be cross.
0. She liked the present really.

**M7 Mrs Peabody**
2. She did not know that he was only wanting the time, and thought that he was a robber. She was worried that she would get attacked and therefore tried to protect herself by offering her bag.
1. So he did not hurt her. OR She is nervous.
0. Because he was a robber.

**P5 Storm**
2. The tree was hit by lightening.
1. The tree was hit by thunder.
0. It fell over.

**M8 Burglar**
2. The burglar thought that the policeman knew about the robbery and was going to arrest him for the crime, even though he was just stopping him to give him his glove. The burglar probably thought he should give himself up to minimise his punishment.
1. He gave himself up to the policeman to make things better.
0. He wanted to admit to the robbery. OR He needed money to feed his children. OR Because he had done the robbery.
Two enemy armies have been at war for a very long time. Each army has won several battles, but now either side could win. The two sides are equally strong. However, the Blue army is stronger than the Yellow army in foot soldiers and tanks. But the Yellow army is stronger than the Blue Army in aeroplanes. On the day of the final battle there is heavy fog over the mountains where the fighting is about to begin. Low clouds hang above the soldiers. By the end of the day the Blue army has won.

During the war, the Red army captures a member of the Blue army. They want him to tell them where his army's tanks are; they know they are either by the sea or in the mountains. They know that the prisoner will not want to tell them, he will want to save his army, and so he will certainly lie to them. The prisoner is very brave and very clever, he will not let them find his tanks. The tanks are really in the mountains. Now when the other side ask him where his tanks are, he says, "They are in the mountains".

Q. Why did the Blue army win?

Q. Why did the prisoner say that?
7. Scoring Criteria for the Cartoons (In order of presentation)

Examples for each score. M= Mentalising Cartoon P= Physical Cartoon.

**M1 Dog and Hose**
3. Dog is barking through the hose and the man does not know where the noise is coming from.
2. Dog is barking into the hose and the man hears a noise coming out OR There is a sound coming out instead of water
1. The dog is barking OR The man hears something.
0. The man is fat

**P1 Children Crossing**
3. The children are walking exactly the same as on the sign.
2. There is a sign for children walking, and they are walking.
1. The children are crossing the road
0. There is something funny about the car

**P2 Loopy Water**
3. The hose pipe is in a loop so the water also comes out in a loop.
2. The water is in a loop.
1. The water misses the flowers.
0. The man looks funny.

**M2 Tarzan**
3. Tarzan thought he was swinging on a vine but he discovers that it is a snake’s tail.
2. Tarzan is swinging on a snake instead of a vine.
1. Tarzan is swinging on a snake.
0. He has got no clothes on.

**P3 Ladder**
3. The men bumped into each other as they passed on the ladder making the lines crooked in the middle
2. The lines were supposed to be straight but the are crooked OR The men are going up and down the ladder in opposite directions
1. The men are painting lines.
0. The ladder will fall over.

**M3 Bull Fighter**
3. The bull does not know that it will bang its head on the iron which the man has hidden behind his cape.
2. The bull will hit its head OR The man has hidden the weight behind his cape.
1. There is an iron weight.
0. The man has no trousers on.
P4 Elephant
3. The elephant has gone to bed and has taken his tusks out like false teeth.
2. The elephant has taken his tusks out.
1. The elephant is in bed.
0. The bed is too small.

M4 Bird and Monster
3. The bird thinks that he has got a worm, but it's really part of a monster.
   OR The monster is angry the bird is pulling on his head.
2. The bird thinks he has a worm.
1. The bird is pulling on the monster's head.
0. There is something underground.

P5 Criminal Chains
3. There is a chain on him instead of the seat-belt.
2. There is a chain.
1. The man has been arrested by the police.
0. The men have different expressions.

M5 Dog and Spanner
3. The man is looking for his spanner but the dog buried it, thinking it was a bone.
2. The dog thinks the spanner is a bone.
1. The man is looking for his tool OR The dog has buried the tool.
0. The man has three hands OR The man is trapped under the car.

P6 Mouse TV
3. The mice have set up chairs to make a cinema.
2. The mice will watch TV
1. The mice have put out little chairs.
0. The man sees the mice.

P7 Skier
3. The man comes down the ski jump so fast he spins up and he lands in the tree.
2. The man lands upside down in the tree.
1. The man is upside down OR The man is skiing in a tree.
0. The other skiers are too small.

M6 Snake and Mouse
3. The two mice eating cheese do not notice the snake disguised in Mickey Mouse ears who wants to eat them.
2. The snake is disguised as/pretending to be a mouse OR The mice do not notice the snake.
1. The mice are eating cheese OR The snake wears funny ears.
0. The snake is too short.
M7 Birdbath
3. The bird thinks it's going to have a bath/drink, but really it's a monster pretending so it can eat the bird when it flies into its mouth.
2. The bird thinks it's a birdbath OR the monster is pretending to be a bird bath.
1. The bird is flying into the bath OR The bird bath has eyes and legs
0. The bird is falling out of the sky

P8 Fishing
3. The man has been waiting so long to catch a fish that a spider has spun its web.
2. The man does not see the spider's web.
1. There is a spiders web on the line.
0. There is no fish on the line.

M8 Picnic
3. The family thought they were eating the picnic in the country but it's really a race course and they don't know they are about to be trampled on.
2. The family don't know it's a race course/ they will be trampled on.
1. The picnic is on a race course.
0. The man is French OR The man has a doll.

P9 Lost and Found
3. The man has lost his head so he has gone to lost and found to get it.
2. The man has lost his head.
1. The man has gone to lost and found OR The man has no head.
0. The man has pulled his coat over his head.

M9 Bird with Headphones
3. The bird listening to his headphones can't hear the tiger and doesn't know he is about to get caught.
2. The bird with headphones is about to get caught OR The bird can't hear the tiger coming.
1. The bird is wearing headphones.
0. The bird is bigger than the house.

P10 Minnie Mouse
3. The toilet seat has ears like Minnie/Mickey Mouse.
2. The toilet has ears.
1. The toilet seat has two lids.
0. Minnie will use the loo.

M10 Bus Stop
3. The man thinks he's in a stick up, but it's really just the other man's guitar.
2. The man thinks there is a gun in his back.
1. The guitar is sticking the man in the back.
0. The man and woman have no eyes.
9. Scoring Criteria for the Eyes

Correct answers in bold

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hate</td>
<td>surprised</td>
<td>kind</td>
</tr>
<tr>
<td>2</td>
<td>unkind</td>
<td>cross</td>
<td>surprised</td>
</tr>
<tr>
<td>3</td>
<td>friendly</td>
<td>sad</td>
<td>surprised</td>
</tr>
<tr>
<td>4</td>
<td>relaxed</td>
<td>upset</td>
<td>surprised</td>
</tr>
<tr>
<td>5</td>
<td>feeling sorry</td>
<td>making somebody do something</td>
<td>joking</td>
</tr>
<tr>
<td>6</td>
<td>hate</td>
<td>unkind</td>
<td>worried</td>
</tr>
<tr>
<td>7</td>
<td>feeling sorry</td>
<td>bored</td>
<td>interested</td>
</tr>
<tr>
<td>8</td>
<td>remembering</td>
<td>happy</td>
<td>friendly</td>
</tr>
<tr>
<td>9</td>
<td>annoyed</td>
<td>hate</td>
<td>surprised</td>
</tr>
<tr>
<td>10</td>
<td>bossy</td>
<td>hoping</td>
<td>angry</td>
</tr>
<tr>
<td>11</td>
<td>confused</td>
<td>joking</td>
<td>sad</td>
</tr>
<tr>
<td>12</td>
<td>thinking about something</td>
<td>upset</td>
<td>excited</td>
</tr>
<tr>
<td>13</td>
<td>happy</td>
<td>thinking about something</td>
<td>excited</td>
</tr>
<tr>
<td>14</td>
<td>not believing</td>
<td>friendly</td>
<td>wanting to play</td>
</tr>
<tr>
<td>15</td>
<td>made up her mind</td>
<td>joking</td>
<td>surprised</td>
</tr>
<tr>
<td>16</td>
<td>angry</td>
<td>friendly</td>
<td>unkind</td>
</tr>
<tr>
<td>17</td>
<td>thinking about something</td>
<td>angry</td>
<td>bossy</td>
</tr>
<tr>
<td>18</td>
<td>angry</td>
<td>daydreaming</td>
<td>sad</td>
</tr>
<tr>
<td>19</td>
<td>kind</td>
<td>surprise</td>
<td>not pleased</td>
</tr>
<tr>
<td>20</td>
<td>interested</td>
<td>joking</td>
<td>relaxed</td>
</tr>
<tr>
<td>21</td>
<td>playful</td>
<td>kind</td>
<td>surprised</td>
</tr>
<tr>
<td>22</td>
<td>surprised</td>
<td>sure about something</td>
<td>joking</td>
</tr>
<tr>
<td>23</td>
<td>serious</td>
<td>ashamed</td>
<td>confused</td>
</tr>
<tr>
<td>24</td>
<td>shy</td>
<td>guilty</td>
<td>daydreaming</td>
</tr>
<tr>
<td>25</td>
<td>joking</td>
<td>relaxed</td>
<td>nervous</td>
</tr>
<tr>
<td>26</td>
<td>ashamed</td>
<td>excited</td>
<td>not believing</td>
</tr>
<tr>
<td>27</td>
<td>disgusted</td>
<td>hate</td>
<td>happy</td>
</tr>
</tbody>
</table>
8. happy

remembering

20. joking

interested

friendly

angry

relaxed

happy