Autism is a disorder characterized by severe impairments in social and communicative development, and by abnormalities in flexible behavior and imaginative thinking (American Psychiatric Association, 1987). In addition, these symptoms often have an onset in the first 3 years of life. The social and communicative abnormalities have recently been conceptualized as the result of abnormalities in the development of a "theory of mind" (Baron-Cohen, 1990, 1991a; Baron-Cohen, Leslie, & Frith, 1985; Baron-Cohen, Tager-Flusberg, and Cohen, 1993), that is, in the ability to appreciate the existence of subjective, mental states (such as beliefs, desires, and intentions) and the explanatory role these can play in making sense of people's behavior.

On tests that tap this ability, at simple levels, 3- to 4-year-old normal children and children without autism but with mental handicap demonstrate remarkable competence, while children with autism (with mental ages higher than this) consistently fail (Baron-Cohen, 1989a, 1989b, 1991b; Baron-Cohen et al., 1985, 1986; Charman & Baron-Cohen, 1992; Leekam & Perner, 1991; Leslie and Frith, 1988; Leslie & Thaiss, 1992; Perner, Frith, Leslie, & Leekam, 1989; Reed & Peterson, 1990). Furthermore, this cognitive deficit correlates with abnormalities in social behavior (Siddons, Happé, Risq, Whyte, & Frith, 1990), and pragmatic competence in communication (Eisenmajer & Prior, 1991).
Despite the consistency of the above results, a question surrounds the percentage of children with autism who can pass tests of a simple theory of mind. This percentage varies from 20-35% in different studies. One possibility is that these children might suffer from a specific delay in the development of a theory of mind. They may develop a theory of mind years after it normally emerges. This notion was first proposed by Baron-Cohen (1985, 1989b) who reported suggestive cross-sectional evidence that even those children with autism who passed first-level (simple) theory of mind tests were nevertheless impaired at the next level up (at the equivalent of a 6- to 7-year-old level).

Baron-Cohen (1991c) elaborated this notion of specific developmental delay as follows: Since both pretend play and joint-attention behaviors are thought to be early precursors in the development of a theory of mind (Baron-Cohen, 1989c; Leslie, 1987), children with autism might initially manifest delay at a very early stage in the development of a theory of mind (e.g., before the equivalent of a 1- to 2-year-old level), and thus fail to show pretend play or joint-attention behaviors (Baron-Cohen, 1987, 1989c; Sigman, Mundy, Sherman, & Ungerer, 1986; Ungerer & Sigman, 1981). With time, some children with autism might reach this point but show their delay at the equivalent stage of a 3- to 4-year-old, thus failing first-order belief attribution tests; while yet others might reach this point but be delayed at the equivalent stage of a 6- to 7-year-old, thus failing second-order belief attribution tests (Baron-Cohen, 1989b). On this model, specific delay would be present from the outset on this developmental pathway. (The question of whether the deficit is one of delay or deviance is discussed further in the debate between Burack, 1992, and Baron-Cohen, 1992.)

In this paper we ask just one question: Using a prospective longitudinal design, how much development of a theory of mind might one expect to see in autism? We chose a 7- to 8-year follow-up period with which to examine this question.

**METHOD**

The 20 children with autism who participated in our 1983 study (published in Baron-Cohen et al., 1985) were followed up for retesting in 1990–1991. Of these, 3 (15%) proved to be untraceable. The remaining 85% of the original subjects who were recruited into the 1990–1991 study comprised 13 males and 4 females, all of whom had been diagnosed as autistic according to established criteria (Rutter, 1978).

The group was assessed for mental age (MA) in 1983 and again in 1990–1991. The tests used were the British Picture Vocabulary Scale.
Autism and Theory of Mind Development

(BPVS; Dunn, Dunn, Whetton, & Pintilie, 1982) and the Leiter International Performance Scale (Arthur, 1952). Details of these scores are presented in Table I. Four subjects did not want to be tested on the Leiter at follow-up, so the mean nonverbal MA given is based on the 13 subjects for whom this data existed at both times. It is clear from Table I that despite a 7-year change in CA, the mean MA of the group only increased by 1 to 2 years.

Design

All but one of the subjects were attending some form of training center or special school. Testing was carried out at these centers except in one case, where the testing was done at home. Each subject was tested by one of the two authors, using the Sally-Anne test (Baron-Cohen et al., 1985: modeled after Wimmer & Perner, 1983). In addition, the Ice-Cream Van test (Baron-Cohen, 1989a: modeled after Perner & Wimmer, 1983) was administered if a subject passed the Sally-Anne test. The methods for administering these tests can be found in the references cited above.

RESULTS

Three subjects passed the Sally-Anne test, the remaining subjects failing this. The Ice-Cream Van test (administered to those 3 subjects who passed the Sally-Anne test) was not passed by any of these subjects. The 3 subjects who passed the Sally-Anne test had a minimum CA of 13 years 3 months, a minimum verbal MA of 8 years 1 month, and a minimum nonverbal MA of 8 years 6 months. These 3 subjects did not differ from

<table>
<thead>
<tr>
<th>Year</th>
<th>CA</th>
<th>Verbal MA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Nonverbal MA&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>M</td>
<td>12.2</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.0</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>6.5–15.8</td>
<td>3.0–7.4</td>
</tr>
<tr>
<td>1990</td>
<td>M</td>
<td>19.8</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>13.1–25.8</td>
<td>2.8–15.2</td>
</tr>
</tbody>
</table>

<sup>a</sup>British Picture Vocabulary Scale (Dunn et al., 1982).
<sup>b</sup>Leiter International Performance Scale (Arthur, 1952), based on n = 13.
the other 14 subjects in terms of either CA, \( t(15) = 0.21 \); nonverbal MA, \( t(11) = 1.14 \) based on \( n = 13 \), or verbal MA, \( t(15) = 1.42 \); all \( p > 0.1 \).

**Comparison of Performance on the Sally-Anne Test Across Time**

Of the 17 subjects tested in the present study, 4 had originally passed the Sally-Anne test in 1983–1984. Of these 4 subjects, 2 passed on retesting, and 2 now failed. In addition, 1 subject who previously failed passed on retesting. This might reflect random fluctuation in test performance; it is not possible to ascertain the causes of this unexpected apparent decline in performance by these 2 subjects. Future studies checking the test–retest reliability of this procedure would be useful to test if this might be an explanation. Of the other 13 subjects, 12 failed on both the original test and the retest, only 1 subject out of these 13 (or 7.7%) showing any improvement with time. Given the small sample, we constructed 99% Confidence Intervals (CI) on this proportion of individuals improving over time, using the formula given below:

\[
\text{estimator} \pm \text{reliability coefficient} \times \text{SE}
\]

where estimator = \( \hat{p} \) ( = 1/13, or 0.077), and where reliability coefficient corresponding to a 99% CI = 0.198, and where \( \text{SE} = \sqrt{\hat{p}(1-\hat{p})/n} \). This generated a 99% CI of 7.7% ± 19.8. In other words, this study suggests that from these data, if repeated sampling were possible, the proportion of reasonably high-functioning subjects with autism that one might expect to improve on this Sally-Anne test over a 7-year period would be between 0 and 27.4%.

Finally, comparing performance on the Sally-Anne test in 1983 and 1990-91 using the Fishers Exact Probability test revealed no significant difference in the proportions passing. These results are summarized in Table II.

**Change in Mental Age Across Time**

Between 1983 and 1990–1991, no significant difference was found in nonverbal MA across time, \( t(12) = 1.91, p > .05 \), based on \( n = 13 \), for whom available data existed at both times, although there was a significant increase in mean verbal MA, \( t(15) = 2.56, p < .05 \). Despite this increase in verbal MA, performance on the Sally-Anne task was poor. This finding raises queries about Prior, Dahlstrom, and Squires’s (1990) report of the role of verbal MA in passing theory of mind tests.
**DISCUSSION**

In this study, development of a theory of mind was assessed in 17 subjects with autism, across a 7- to 8-year time interval. In the group as a whole there was no statistical change (23.5% passing the Sally-Anne test of understanding first-order false belief in 1983, and 17.6% passing it in 1990–1991). In the group as a whole, performance was largely unchanged: 82.4% of the group showing identical performance on both occasions. One subject who had previously failed the Sally-Anne test now passed, and 99% Confidence Intervals on the proportion expected to show improvement, given this data, were 0–27%. Two subjects who had passed initially now failed. Of the subjects who passed the Sally-Anne test on follow-up, none passed the Ice-Cream Van test (of understanding second-order false-belief), and this concurs with results from Baron-Cohen (1989a). Whether there is a ceiling on development of a theory of mind in the majority of children with autism somewhere between the equivalent of a 3- to 6-year level is a real possibility, although results from other studies (Bowler, 1989; Happé, 1991) suggest that among higher functioning adults with autism or Asperger syndrome, there may be development of a theory of mind even to the 6- to 7-year level.

In general, outcome after 7–8 years was not accompanied by significant improvement in nonverbal MA, nor by any dramatic improvement in social independence, although language levels had significantly increased. This relative consistency in functioning over time is in line with other follow-up studies (DeMyer et al., 1973; Gillberg, 1991). The lack of increase in nonverbal MA could be taken to indicate an overall decrease in IQ, given that IQ = MA/CA. If this is so, this would be inconsistent with previous studies showing the stability of IQ over time, in autism (Rutter, 1978). This question merits further investigation.

From this study, we conclude that for the majority of people with autism (approximately 60–70%), there may be little development of a

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Fail</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

**Table II. Comparison of Performance on the Sally-Anne Test Across Time**
theory of mind. For a minority (approximately 20–30%), development may extend to the equivalent of a 3- to 4-year-old level by the teenage years. Finally, from other studies, it is clear that some of these individuals may reach the equivalent of a 6- to 7-year-old level, but often only by adulthood. Whether such individuals are showing late development of the same processes as are used normally, or are arriving at the same point via an alternative cognitive compensatory route (Frith, Morton, & Leslie, 1991), is an open question. Given the small sample, we feel it only proper to introduce a note of caution in interpreting these results, and publish them at this stage in the hope that they might form part of a more meaningful data bank as other research groups carry out similar follow-up studies.

REFERENCES


