Deception and Sabotage in Autistic, Retarded and Normal Children

Beate Sodian* and Uta Frith†

Abstract—We investigated autistic, mentally retarded, and normal children’s ability to deceive or obstruct an opponent. When required to tell a lie (saying that a box was locked) autistic children performed significantly worse than their controls, taking into account mental age. However, they readily prevented a competitor from gaining a reward by physical manipulation (locking a box). Their success on sabotage demonstrated that their failure on deception was not due to an inability to understand the task. Performance on deception was predicted by performance on a false belief attribution task. The present findings confirm that autistic children have a specific deficit in understanding and manipulating beliefs.

Keywords: Autism, theory of mind, deception

Introduction

In predicting and explaining other people’s actions we seem to use a common-sense psychological theory, interpreting human behaviour in terms of beliefs and intentions. It is now fairly well established that by age 4 years normal children have already acquired the most important principles of this common-sense theory of mind (see Astington, Harris & Olson, 1988; Miller & Aloise, 1989; for reviews). It has also been established that autistic children lack this fundamental ability: even when their mental age is 4 years and above (see Leslie & Frith, 1990, for a review), autistic children fail tasks testing their understanding of others’ ignorance and false belief (Baron-Cohen, Leslie & Frith, 1985; Perner, Frith, Leslie & Leekam, 1989), they do not understand event sequences in psychological-intentional terms (Baron-Cohen, Leslie & Frith, 1986), and they fail to distinguish between mental and physical entities, and between appearance and reality (Baron-Cohen, 1989). All these tasks are solved by normal 4-year olds (some by even younger children) and by non-autistic retarded children of comparable mental age. Thus, the conclusion can be drawn that a failure to impute mental states despite adequate intellectual capacity is specific to autism. The cognitive deficit that leads to this failure can explain the severe social, affective,

Accepted manuscript received 20 March 1991

*Universität München, Institut für empirische Pädagogik, Leopoldstr. 13 8 München 40, Germany.
†MRC Cognitive Development Unit, 17 Gordon Street, London WC1H 0AH, U.K.
Requests for reprints to: Dr Uta Frith, MRC Cognitive Development Unit, 17 Gordon Street, London WC1H 0AH, U.K.
and communicative impairments that are the critical behavioural symptoms observed in autistic children (Frith, 1989).

A theory of mind guides the intentional manipulation of other people’s knowledge and beliefs. If children are unable to represent a belief that deviates from a known state of affairs, they should be unable to induce deliberately such a belief in another person. Recent studies by Peskin (1989), Russell, Mauthner, Sharpe and Tidswell (1991), and Sodian (1991) indicate that the ability to deceive an opponent emerges at around the age of 4 years in normal children, as would be expected from previous research on false belief representation (Wimmer & Perner, 1983; Perner, Leekam & Wimmer, 1987). While these studies of deception found that 3-year olds failed to deceive in a variety of tasks even after practice and under very conducive conditions, Chandler, Fritz and Hala (1989) report the use of deceptive ploys in a hiding-game context (erasing footprints and laying false trails) even in 2½-year olds. However, a replication study by Sodian, Taylor, Harris and Perner (1991) showed that 3-year olds used these ploys indiscriminately for a competitor as well as for a cooperator, thus indicating that very young children can be led to use deceptive techniques without understanding their effect.

Like young normal children, autistic children seem to be capable of manipulating other people’s behaviour (e.g. by using instrumental gestures); they do not, however, use expressive gestures which are normally used to manipulate other persons’ mental states (Attwood, Frith & Hermelin, 1988). Anecdotal evidence suggests that autistic children do not lie and do not understand deception. Indeed, parents and teachers are greatly concerned about the practical consequences of this aspect of social naivety in autistic adults who live semi-independent lives. Autistic children’s ability to deceive has, however, not been studied systematically with the exception of a study by Russell et al. (1991). In this study normal children aged 3 and 4, Down’s syndrome children and autistic children with a mental age of around 4 years or above were compared. The scenario involved the child and a competitive experimenter, each of whom could win sweets that were hidden in one of two boxes. Each time the child had to point to one of the boxes in order to “tell” the competitive experimenter where to look for a sweet. Whenever the experimenter found the sweet, he/she kept it; if not, then the child got it. This game was established during 15 trials. Trial 16 was the test trial. Here two new boxes were used. Now, the child could see which box contained the sweet, through a window in the box, but the competitor could not. Most 4-year olds and most Down’s syndrome children immediately pointed to the empty box, often giggling, so that they and not the experimenter got the sweet. In contrast, most 3-year olds and nine out of 11 autistic children pointed to the full box and persisted in doing so for another 19 trials.

In the present study we wanted to investigate just how profound and specific autistic children’s failure to deceive is. For this purpose, we included control groups of young normal and mentally retarded children. The mentally retarded children in this study differed from the control groups used in previous studies in that we did not restrict them to Down’s syndrome. Down’s syndrome children are often described as being particularly socially “friendly” and interested. Little is known about the ability to understand mental states in heterogeneous groups of handicapped children, most of whom do not have a record of particularly good social skills. Nevertheless, we chose
such a group in the hope of providing a particularly stringent test of the claim that
the failure to develop a concept of belief is specific to autism.

We used two tasks. One was modelled after Woodruff and Premack’s (1979) study
of deception in the chimpanzee and required deceptive pointing (similar to Russell
et al.’s, 1991, study). The other task required telling a lie (telling a competitor that
a box containing a reward was locked when in reality it was open). Both tasks were
embedded in the context of a hiding game in which children were competing for a
reward with a “nasty” puppet. There was also a condition in which a “nice” puppet
helped children win rewards. This served as a cooperative control condition. Sodian
(1991) found that performance on these two tasks was highly correlated in normal
3-4½-year-old children, indicating that once children have acquired the concept of
belief they easily employ it in a whole range of social manipulation tasks. This
interpretation is also suggested by the remarkable consistency between the findings
of studies that used quite different deception tasks (Peskin, 1989; Russell et al., 1991;
LaFreniere, 1988; Sodian, 1991).

Most deception tasks can, however, also be solved by behavioural strategies, without
representing beliefs. After all, a stereotyped reaction can be evoked through a well
established behavioural routine. This argument was in fact used by Dennett (1978)
when discussing Premack and Woodruff’s (1978) study. It is unlikely that normal
4-year olds solve deception tasks in this way, given that they have been shown to
succeed in novel tasks on the first trial (Russell et al., 1991; Sodian, 1991), and that
performance on deception tasks correlates highly with the ability to attribute a false
belief to a story character (Russell et al., 1991). It is possible, however, that in older
and cognitively advanced autistic children some routines for manipulating other
people’s behaviour (e.g. pointing to a place to make a person go there) have become
well established and can be employed as if to deceive, without any proper understanding
of the effects of these behaviours on other persons’ beliefs. It is therefore important
to bear in mind that autistic children’s successful responses in deception and belief
attribution tasks do not necessarily indicate the presence of a belief concept.

If we want to argue that autistic children’s presumed difficulty with deception is
due to a specific failure to acquire a concept of belief, we have to rule out the possibility
that they fail deception tasks for other reasons, such as a general inability to form
“anti-plans” (i.e. plans about how to prevent a competitor from executing his plan
and reaching his goal) or such as a lack of interest in competitive games. Sodian (1991)
showed that normal 3-year olds failed deception although they passed parallel
sabotage tasks that required anti-planning but did not require the manipulation of
beliefs (e.g. locking a box to prevent a competitor from gaining a reward). Most normal
children even mastered a relatively complex sabotage task, in which they were required
not only to lock a reward-containing box for a competitive nasty puppet, but also
to lock an empty box to make sure the nice puppet immediately found the reward.

For children to demonstrate competence at sabotage they have at least to be capable
of physically preventing a competitor from attaining a reward. In our tasks this could
be done by locking a reward-containing box. To rule out the possibility that locking
a box is merely a form of stereotyped behaviour, the children also had to leave open
the box for a cooperative partner. Only then did they reach the minimum criterion
for successful sabotage. Locking an empty box for a cooperative partner provided
an additional, more conservative measure of children’s competence at anticipating and manipulating other people’s behaviour in tasks that do not require the representation of beliefs.

In sum, we predicted from previous research that autistic children would fail to actively deceive an opponent even under very conducive task conditions. If this presumed inability is specific to autism, then autistic children should perform significantly worse on deception tasks than normal or retarded children of comparable mental age. Furthermore, if the presumed deficit is specific to the manipulation of beliefs, then autistics (as well as young normal and retarded children) should be able to succeed on parallel tasks requiring them to hinder an opponent from attaining his goal while not requiring them to represent the opponent’s beliefs (sabotage tasks). These predictions were tested in two deception tasks that differ in the kind of deceptive action required (telling a lie vs deceptive pointing) and two parallel sabotage tasks that differ in the complexity of the behavioural strategy required.

Method

Subjects

Nineteen autistic children, eight girls and 11 boys, diagnosed by established criteria (APA, DSM-III-R, 1987), were tested in various special schools in England and through a parent organization in Munich. As shown in Table 1 their chronological ages ranged from 6 to 19 and their mental ages, assessed by the British Picture Vocabulary Scales, ranged from 4 to 12; five of the 19 autistic children were relatively able, with a verbal mental age above 7:6. Twenty-nine mentally retarded (i.e. learning disabled) children, 17 boys and 12 girls, were tested in two special schools in Munich. Their chronological ages ranged from 10:0 to 16:7, and their mental ages, assessed by the verbal part of the HAWIVA (the German form of the Wechsler Intelligence Test for preschool children), and the British Picture Vocabulary Scales (BPVS), ranged from 2:10 to 7:0; 18 of these children had a verbal mental age above 4. These children’s general learning disability was moderate to severe, and was of unknown aetiology. Thirty-nine normal children, 19 boys and 20 girls, were tested in day-care centres in Munich. Nineteen of these children ranged from 3:0 to 4:0, and 20 were between 4:1 and 5:2 years old. The normal children were also tested on HAWIVA and BPVS and their mental ages were found to be roughly in accord with their chronological ages. Table 1 shows the grouping according to verbal mental age that we adopted for comparisons between diagnostic groups.

Materials

Two large hand-puppets, a white seal and a fierce-looking wolf, were used as interaction partners. Two pink toy money boxes that were made distinctive with a blue and a red ribbon were used as hiding places. The boxes could be locked with a small padlock. Sweets (Smarties, M & M’s), were used as target items. A doll was used as a stooge for a questioning procedure following the sabotage and deception tasks. For the standard false belief attribution task, a box, a purse, and a basket were used as hiding places, and a coin as a target item.

Procedure

In both the “one box” and the “two boxes” tasks, the experimenter started by introducing the puppets. The seal was introduced as the nice smartie friend who gives smarties to the children, and the wolf was introduced as the nasty smartie eater who takes smarties away. To demonstrate the puppets’ behaviours, one of the toy boxes was placed on the table, a smartie was put into the box, and the smartie friend appeared holding another smartie in his paws. He went up to the box, opened it, discovered the
Table 1. Subjects grouped according to diagnosis and verbal mental age*

<table>
<thead>
<tr>
<th>MA band</th>
<th>Autistic</th>
<th>Mentally retarded</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 11</td>
<td></td>
<td>n = 19</td>
</tr>
<tr>
<td>Very low</td>
<td>CA 10:0–15:0</td>
<td>CA 3:1–4:0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 11:6</td>
<td>mean 3:6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA lower than 4:0 (below test norms)</td>
<td>(younger group)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>n = 8</td>
<td>n = 9</td>
<td>n = 20</td>
</tr>
<tr>
<td></td>
<td>CA 6:0–17:9</td>
<td>CA 10:0–14:1</td>
<td>CA 4:1–5:2</td>
</tr>
<tr>
<td></td>
<td>mean 12:7</td>
<td>mean 12:5</td>
<td>mean 4:7</td>
</tr>
<tr>
<td></td>
<td>MA 4:0–5:5</td>
<td>MA 4:3–5:3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 5:0</td>
<td>mean 4:6</td>
<td>(older group)</td>
</tr>
<tr>
<td>Middle</td>
<td>n = 6</td>
<td>n = 9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA 8:9–19:2</td>
<td>CA 10:7–16:7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 14:8</td>
<td>mean 13:2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA 5:6–6:3</td>
<td>MA 5:6–7:0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 5:11</td>
<td>mean 6:3</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>n = 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA 12:3–16:6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 14:5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA 7:6–12:0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean 8:9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Verbal MA was assessed by the British Picture Vocabulary Scale, Long Form (BPVS) and the German version of the WPPSI Vocabulary subtest HAWIVA.*

smartie, and put the second smartie into the box, offering the contents of the box to the child. The experimenter explained that the smartie friend gives the child another smartie whenever he finds one. Next the smartie eater appeared, went up to the box, opened it, and took the contents out, eating them up. The experimenter explained that the nasty smartie eater takes a smartie away whenever he finds one and eats it up. The child was then asked to identify the nice smartie friend and the nasty smartie eater, and the experimenter introduced the task as a game in which the child could win many smarties. She said: “To win lots of smarties, you have to help the nice smartie friend to find them,” (pointing at the seal), “but don’t let the nasty smartie eater get them” (pointing at the wolf).

One box task. The two conditions are illustrated in Figs 1 and 2. The sabotage condition was introduced by demonstrating that the box could be locked with a padlock. Then the child was instructed to hide a smartie in the box. The experimenter then produced the friend/eater from under the table. The friend said: “I am the nice smartie friend. When I find a smartie, I give you another one.” The experimenter proceeded: “Help him find the smartie” and asked the test question: “Do you want to lock the box or do you want to leave it open?” On the competitive trial, the eater said: “I am the nasty smartie eater. When I find a smartie, I eat it up.” The experimenter proceeded: “Don’t let the nasty smartie eater find it” and asked the test question: “Do you want to lock the box or do you want to leave it open?”

In the deceptive condition, there was no key to lock the box. The hiding procedure was the same as in the sabotage condition. When the eater/friend appeared, the experimenter held the puppet far off and the puppet said: “I wonder whether this box is locked or open. I can’t see it from here.” (The lock was always on the side facing the child.) “If it’s locked I won’t bother to make the long walk.” The puppets then proceeded introducing themselves in the same way as in the sabotage condition, and the experimenter instructed the child to help or not help the puppets in the same way as described above. The smartie eater/friend then asked the child: “Is this box locked or is it open?” and the
experimenter proceeded with the test question: "What do you want to say? Do you want to say it is locked or do you want to say it is open? Think carefully. Help the friend to find the smartie./Don't let the eater find the smartie." If the child lied to the eater, she could keep the smartie, but if she told the truth, he took it away. If she told the friend the truth she was given an additional smartie reward.

*Post-test probe for explanation of motives.* After the child had completed two pairs of trials in both the deception and the sabotage conditions, the experimenter introduced a doll, "Monica", who received the same tasks as the child (one cooperative and one competitive trial in each of the sabotage and deceit
Fig. 2. Deception (drawing by Axel Scheffler, reproduced with permission).

conditions). The doll locked the box for the competitor, left it open for the cooperator, said it was locked to the competitor and said it was open to the cooperator. After each of her actions/utterances, the child was asked: "Why did she do this?/Why did she say this?" The child's answers were followed up with questions attempting to explore the child's understanding of the motives for and the effects of sabotage and deception.

Two boxes task. The sabotage condition was introduced by demonstrating that each of the two boxes could be locked with a padlock. Then the padlock was placed in the front of the child in the middle
between the boxes. The child was given a choice in which box to hide a smartie. The eater/friend appeared, introduced in the same way as in the "one box" task. The experimenter instructed the child to help the friend/not to let the eater find it, and asked the test question: "What do you want to do? Do you want to lock this box or do you want to lock that box?" (indicating the full vs empty box in counterbalanced order).

In the deception condition, the procedure was the same except that there was no padlock, and the eater/friend asked the child: "Where is the smartie?" whereupon the experimenter asked the test question: "Where do you want to point? Do you want to point to this box (full) or do you want to point to that box (empty)? Think carefully. Help the friend to find the smartie; don't let the eater find the smartie."

Standard false belief task. For the standard false belief attribution task, modelled on Leslie and Frith (1988), a purse, a basket and a box were placed on the table. Experimenter 2 produced a coin and said: "Look, I have a coin. I'll put it into the basket. I have to go outside for a minute; I have to fetch a handkerchief. Will you stay here and watch my coin?" When E2 had left the room, Experimenter 1 asked the child: "Where did E2 put the coin?" She then said: "Look, what I am going to do now: I take the coin out of the basket and put it somewhere else." Sometimes the child made the transfer; sometimes the experimenter. She then asked a series of test questions: False belief 1: "When E2 comes back, where will she look for the coin?"; Memory control 1: "Where is the coin now?"; Memory control 2: "Where did E2 hide the coin?"; False belief 2: "Where does E2 think the coin is?".

Design

It was intended that each child should receive all these tasks. This took two sessions, no more than a week apart. In the normal and retarded groups, however, some children could not be tested twice and thus only received one task. Five normal children did not receive the "two boxes" and six did not receive the "one box" task. Two retarded children did not receive the "two boxes" and one did not receive the "one box" task. Half the children received the "one box" task first, the other half started with the "two boxes" task. The sabotage and deception conditions each included two pairs of trials, a pair consisting of one cooperative and one competitive trial. Half the children in each group started with a sabotage pair of trials, followed by a deception pair, and another sabotage and deception pair; the other half started with a deception pair of trials. The order of cooperative and competitive trials was counterbalanced across subjects. At the end of the "one box" task a standardized questioning procedure was used in which the child was asked to explain why a doll either deceived, or physically hindered, the competitor and why she helped the cooperator. The standard false belief task was added to either one of the two sessions.

The response patterns for passing or failing on deception and sabotage tasks were scored according to the design shown in Table 2. In order to be credited with passing a trial in a given condition, the child had to perform correctly on a complete pair of trials, i.e. with both competitor and cooperator. A score of 1 was given for correct performance on one complete pair of trials with competitor and cooperator. A score of 2 means 2 pairs correct (maximum for each of four conditions). A score of 0 means no single pair was correct, but single occasions with either competitor or cooperator may have been correctly performed.

<table>
<thead>
<tr>
<th>Table 2. Design of deception/sabotage tasks and correct responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>One box</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Two boxes</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Results

The performance of the three groups on all conditions was subjected to analysis of variance with MA as covariate. The MA adjusted means are shown in Fig. 3.

![Graph showing performance of three groups under different conditions.](image)

Fig. 3. MA adjusted means by conditions. N = normal, MR = mentally retarded, AUT = autistic; FB = false belief; SAB. 1, DEC. 1 = one box sabotage and deception; SAB. 2, DEC. 2 = two box sabotage and deception.

An overall ANCOVA contrasting one box and two box tasks as well as deception and sabotage resulted in a significant interaction of groups by tasks by conditions \(F = 3.49, \text{df}(2,69), p < .04\). No other term reached significance. Post-hoc t-tests revealed that this was due to the autistic group differing both from the normal \((p < .01)\) and from the learning disabled group \((p < .02)\) in terms of their extremely poor performance on deception compared to sabotage on the one box task.

This single significant interaction was explored further by a series of ANCOVAs for each of the experimental conditions. On the one box deception task a very specific failure was demonstrated by the autistic children \(F = 4.92, \text{df}(2,69), p < .01\). Post-hoc tests showed a clear-cut difference between autistic and both normal and learning disabled children \((p < .01)\) who did not differ from each other. In contrast, autistic children’s performance on the simple sabotage condition was just as good as that of the controls.

As implied by the outcome of the overall ANCOVA, the results for the two box tasks did not isolate a specific impairment of the autistic group. Rather, on both deception and sabotage conditions the two handicapped samples resembled each other and both showed poorer performance than the normal children. \(F = 3.49, \text{df}(2,69), p < .05; \text{for sabotage } F = 9.56, \text{df}(2,69), p < .001\). Post-hoc comparisons of the autistic vs normal groups were significant at \(p < .01\), and of the learning disabled vs normal groups, marginally at \(p < .06\) for deception, and \(p < .001\) for sabotage.

False belief attribution

After controlling for MA through ANCOVA, performance on our standard false belief task showed a significant group difference \(F = 13.33, \text{df}(2,69), p < .0001\).
Post-hoc tests revealed that the autistic group had significantly poorer performance on the false belief task than either the normal ($p < .0001$) or the learning disabled group ($p < .02$). The latter also differed from the normal group ($p < .001$). While the learning disabled children found the false belief attribution task particularly difficult, the autistic children found it more difficult still. All children answered the two memory control questions correctly. Confirming earlier findings, performance increased between the ages of 3 and 4 for normal children, from six out of 12 to 14 out of 16 (Fisher’s exact test, $p = .04$). However, only eight out of 16 learning disabled children with a mental age of 4 and above succeeded on the standard false belief task, as well as four of the 14 autistic children in the same MA band.

Since we predicted a close relationship between the ability to pass false belief and deception tasks we added false belief performance as second covariate for a further series of analyses. Our prediction was confirmed since we found that covarying out false belief now removed any group differences on the two deception conditions. In both cases the effect of partialling out false belief was more powerful ($p < .0001$) than that of MA ($p < .04$ for one box and $p < .1$ for two box). Likewise, in stepwise multiple regression analysis false belief performance predicted performance on lying [$F($df 1, 69) = 19.7, $p < .001$] and deceptive pointing [$F($df 1, 69) = 15.9, $p < .001$], while neither group membership nor MA contributed anything further. Thus we can assume that both tasks tap the same underlying cognitive ability independent of mental age.

While simple sabotage did not differentiate the groups in any case, complex (two box) sabotage showed an unexpected effect. Here, after MA and false belief were covaried out, the group difference persisted [$F = 3.19$, df(2,68), $p < .05$] with normal children showing superior performance to the handicapped groups ($p < .05$). There is then in this particular condition an unknown factor that penalizes the handicapped children over and above any general intellectual impairment and regardless of theory of mind related capacity.

False belief was not significantly more difficult for autistic children than either of the deception tasks. However, the false belief task proved to be more difficult than the deception task for mentally retarded children ($p = .002$), whereas the opposite pattern emerged for normal children ($p = .004$).

Justification for the doll’s responses

In general, the children’s justifications for the doll’s actions were disappointingly sparse. In many cases the spontaneous explanations did not allow a distinction between an understanding of the effects of deception on an actor’s behaviour and on an actor’s belief. There was a tendency for all children to justify both deception and sabotage in functional terms, typically, “so that the wolf doesn’t get the sweet”.

Nevertheless, 50% of the normal children who were successful on the deception tasks made reference to the partner’s belief in their explanations. For instance, “because the wolf thinks there’s no sweet in the box”. The successful children in the mentally retarded and autistic groups almost always used behavioural justifications, e.g. “because it’s empty”, “so that the wolf doesn’t eat it up”. On the other hand, use of the terms, “thinks, lies, jokes, wants” (not necessarily correct use) could often
be elicited by questions. This was also the case for six out of 10 autistic children who themselves had failed at both deception tasks. There was therefore no close link to be made between successful performance and successful explanation of performances with reference to mental state terms.

Discussion

Autistic children’s difficulty with belief representation has been amply documented. Our study demonstrates that many autistic children not only fail to infer other people's false beliefs but they also fail to manipulate other’s beliefs in quite simple and conducive situations. This failure was particularly striking on the "one box" task, in which children had to tell a simple lie to prevent an opponent from winning a reward. This task was easily passed by almost all normal 4-year olds and by all retarded children with a mental age of about 5 years. In contrast, almost all autistic children with a mental age between 4 and 5 years failed this task, and even in the group with a verbal mental age between 7 and 12 only 60% succeeded.

Given anecdotal reports on autistic children's and adolescents' failure to lie, this result was expected. The main question motivating our experimental study was whether this deficit is specific to the manipulation of beliefs or whether it reflects a more general cognitive or motivational problem in social interaction. We therefore introduced a control task that was similar to the experimental task in requiring children to prevent an opponent from attaining a reward but differed in the means that could be employed to attain this goal. The results showed that autistic children happily prevented a competitor from getting a sweet by locking the box in which it was hidden. More importantly, they did not apply this strategy indiscriminately, but left the box open for a cooperative partner. This is remarkable as one might have expected autistic children to show stereotyped behaviour (i.e. lock the box regardless of experimental condition). Thus, when physical means were available, autistic children helped the cooperator and hindered the competitor. They failed in this objective only when physical means were not available. This result is fully predicted from the metarepresentational account of autism. This account, which has been related in some detail to the clinical picture of autism by Frith (1989), postulates a specific deficit in the representation of mental states such as beliefs, not a more general impairment in social interactive skills.

One criticism of our present experiment is that the language demands of the lying task might have made this task more difficult than the parallel sabotage task and that this difference in difficulty would account for the results regardless of its metarepresentational requirements. Children had to understand fairly elaborate instructions, such as "do you want to say it is locked or do you want to say it is open?" At the same time one would have to assume that these instructions were easily understood by the mentally retarded children in order to account for their good performance. While we cannot discount this sort of criticism from our study, we can point to a recent experiment which poses minimal language demands but also attests to autistic children’s specific difficulty with deception. Baron-Cohen (1990) explored autistic children's performance in a naturalistic guessing game, the hide-a-penny game
that has been previously employed with normal children (Gratch, 1964; De Vries, 1970). This game is essentially explained and conducted non-verbally. The results revealed that autistic children enjoyed the game as a game of “object occlusion” (keeping things out of sight), but failed to employ strategies of information occlusion (not giving clues to the hiding place). Normal 4-year olds and mentally retarded children of comparable mental age proficiently employed such information occlusion strategies. These results indicate that autistic children fail to engage even in the simplest forms of information manipulation, and that this failure cannot be attributed to language comprehension problems.

Consistent with Russell et al.’s (1991) results, we also found that most autistic children with a mental age of about 4–7 years failed to employ a deceptive pointing strategy (pointing to an empty box) to mislead an opponent. However, their failure on this task was not as pervasive as with one box deception and not as pervasive as we expected from Russell et al.’s (1991) findings on a very similar task. Furthermore, they performed no worse than their mentally retarded controls. Again, we investigated the specificity of this deficit by introducing a parallel sabotage task that required physical prevention (choosing which box to lock as opposed to which box to point to). Almost all autistic children correctly locked the full box when it was the competitor’s turn to come and try to get the smartie. To make this sabotage task parallel to the deceptive pointing task, it was necessary to introduce a rather complicated condition: to help the cooperator, the child had to lock the empty box and leave the full box open. Replicating results from a previous study (Sodian, 1991), normal 3-year olds were significantly more successful on this complex sabotage task than on the deceptive pointing task. The same tendency was found in autistic children, but the difference was not significant. Thus, there was no unequivocal evidence that autistic children can employ very sophisticated behavioural manipulation strategies while they still fail extremely simple belief manipulation tasks. On the other hand, there was evidence of special problems with complex sabotage for both our handicapped groups.

A possible explanation for this finding is that both the cooperative part of the sabotage task and the competitive part of the deception task required children to focus on and operate upon an empty box (either locking it or pointing to it), while at the same time inhibiting a response towards a full box, where the reward was hidden. Thus, autistic children’s difficulty with the two box task and also that of the learning disabled children may reflect a more general problem in the executive control of behaviour (see Russell et al., 1991, for a proposal along these lines). Autistic and retarded children may be irresistibly drawn towards the location where a salient object really is, unable to inhibit truthful pointing or a verbal description of a state of affairs that corresponds to reality. Russell et al. (1991) argue that such a deficit in the executive control of behaviour could account for autistic children’s failure on the standard false belief task where they are drawn to point to the actual location of the object (and not where the other person must still think it is). This proposal clearly warrants further exploration. Note, however, that it is not consistent with our evidence from the one box task. Autistic children’s good performance on the simple sabotage task in our study is evidence for an ability to control a behavioural tendency when no belief manipulation is involved.

While there was no evidence from this study that autistic children’s ability to
represent beliefs was underestimated in belief attribution tasks, there was evidence supporting this assumption in the retarded children who, as a group, performed significantly better on deception than on belief attribution. The retarded children performed relatively poorly on our standard false belief task, also compared to Down's syndrome children studied in previous experiments (Baron-Cohen et al., 1985, 1986). Generally, the tasks that were mastered by normal 4-year olds were only mastered by retarded children in their teens, with a mental age above 5 years. Thus, it is possible that not only autistic children but also some subgroups of mentally retarded children acquire a concept of belief with a marked delay. This possibility is also suggested by Prior, Dahlstrom and Squires (1990). Note, however, that mentally retarded children did perform significantly better than autistic children when mental age was partialled out on both the false belief and simple deception tasks. This is consistent with other evidence indicating that the impairment found in autistic children cannot be attributed to retardation *per se*. Our study suggests that simple deception tasks might be particularly sensitive indicators of the deficits that we presume to be specific to autism.

The inclusion of a subgroup of very able autistic children in our study made it possible to address the question of whether there is developmental progress in autistic children's acquisition of a concept of belief. Do autistic adolescents eventually come to understand belief albeit with a gross delay? We found that there was indeed some developmental progress with mental age. On the deceptive pointing task, in particular, about 80% of the highest mental age group (7-12 years) succeeded. Caution is warranted, however, in interpreting this finding as evidence for belief understanding, since behavioural strategies that have the effect of deceiving an opponent may be employed without actual deceptive intention on the part of the child. Four autistic children who showed perfect performance on both the lying and the deceptive pointing tasks also passed the belief attribution task. These are the most likely candidates to have developed a genuine understanding of belief, possibly with a marked delay. However, the quality of the verbal justifications of these children does not give grounds for optimism. One very articulate 18-year old boy explained the doll's motive for performing sabotage very precisely "to stop the wolf from opening the box", but when asked to explain the doll's lie, all he said was "because", and left it at that even when questioned further. Likewise, a 19-year old girl said in relation to sabotage "so the thief can't get it, 'cos if he does he'll see there is a sweet and take it and then she'll lose a point." In relation to lying she simply said "'cos there is a sweet in the box and the wolf is coming". Only one outstandingly able 12-year old girl gave an explanation that referred to the thief's false belief. "The thief thinks it's in the empty box. But it's not there."

It would clearly be important to investigate the successful children further and to establish if there is evidence for a theory of mind in their everyday behaviour or whether they have developed relatively sophisticated behavioural strategies to compensate for their deficit. This question has practical implications, for instance, for forensic evidence involving autistic individuals, and particularly for efforts to teach and train social understanding in autistic children. Parents and teachers in fact try to teach autistic children to "understand" jokes, lies and others' beliefs. It would be extremely interesting to see whether such efforts eventually lead to the acquisition of a concept
of belief that can be applied flexibly to a variety of situations, or, whether all that is acquired is a certain number of behavioural routines.

Acknowledgements—The cooperation of the authors was made possible by a travel grant from the Deutsche Forschungsgemeinschaft to the first author (SO 213/2-1). We are grateful to Aurelia Hoelker and Petra Schlaba, Munich, and Fran Siddons and Francesca Happe, London, for their assistance with data collection, and to Chris Frith for statistical advice. We thank the following schools for their kind permission to carry out the experiment: Broomhayes School, Devon; Heathlands School and Springhallow School, London; Heilpaedagogisches Zentrum Augustinum; Montessori Schule; Kleinkindergruppe St. Ignatius, Munich, and the Munich association of autistic children’s parents.

References


---

**Appendix**

**Table A1. One box task. Number of children who were correct on two, one or no pair of trials on the sabotage and deception tasks**

<table>
<thead>
<tr>
<th>MA band</th>
<th>Sabotage (simple)</th>
<th>Deception (lying)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Autistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Middle</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Mentally retard</td>
<td>Very low</td>
<td>4</td>
</tr>
<tr>
<td>Low</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Middle</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Normal</td>
<td>Very low</td>
<td>10</td>
</tr>
<tr>
<td>Low</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

*a For definition see Table 1.

**Table A2. Two boxes task. Number of children who were correct on two, one or no pair of trials on the sabotage and deception tasks**

<table>
<thead>
<tr>
<th>MA band</th>
<th>Sabotage (complex)</th>
<th>Deception (pointing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Autistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Middle</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>High</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mentally retard</td>
<td>Very low</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Middle</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Normal</td>
<td>Very low</td>
<td>9</td>
</tr>
<tr>
<td>Low</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

*a For definition see Table 1.*