Understanding the Mind or Predicting Signal-Dependent Action? Performance of Children with and without Autism on Analogues of the False Belief Task.

Dermot M Bowler
City University, London

Jackie Briskman
Institute of Psychiatry, London

Nicole Gurvidi
City University, London

and

Miriam Fornells-Ambrojo
St Thomas' Hospital Medical School
London

Running Head: Signal understanding in autism.

Address Correspondence to: Dermot M Bowler
Department of Psychology
City University, London
Northampton Square
London EC1V 0HB
U.K.

e-mail d.m.bowler@city.ac.uk

Abstract

To test the claim that correct performance on unexpected transfer false belief tasks specifically involves mental state understanding, two experiments were carried out with children with autism, intellectual disabilities and typical development. In both experiments, children were given a standard unexpected transfer false belief task and a mental-state-free, mechanical analogue task in which participants had to predict the destination of a train based on true or false signal information. In both experiments, performance on the mechanical task was found to correlate with that of the false belief task for all groups of children. Logistic regression showed that performance on the mechanical analogue significantly predicted false belief task performance even after accounting for the effects of verbal mental age. The findings are discussed in relation to possible common mechanisms underlying correct performance on the two tasks.

Understanding the Mind or Predicting Signal-Dependent Action? Performance of Children with and without Autism on Analogues of the False Belief Task

It is now well established that three year old typically developing children fail unexpected transfer tests of false belief such as that used by Wimmer and Perner (1983). In these tasks, children are asked to predict the behaviour of a protagonist who thinks an object is in one location when it really is in a different one. Despite the consensus on the findings, there is still considerable debate over why children should experience such difficulty.

Theorists such as Perner (1991) argue that young children find it hard to understand the representational nature of mental states. That is, pre-school children, although they can understand that perceptual access to an event results in an individual's becoming aware of that event, cannot understand that a person's knowledge can become out-dated by changes in events that were not witnessed by the individual concerned. Perner's evidence shows that it is not until their fifth year that children become aware of the fact that other people's mental states are representational, i.e. that they can bear true <u>or</u> false relation to reality. This development by four-year-old children of an understanding of the representational nature of mental states such as beliefs is often referred to as their acquiring a 'theory of mind'. This account was by Zaitchik's (1990) findings on the false photograph task. This task involves an experimenter placing an object in one location and taking a photograph of it. The object is then moved to another location, and the child is asked ‘In the picture, where is the object?’. Zaitchik found that similar proportions of children passed this task as passed the false photograph task and concluded that what developed during the child’s fifth year was an understanding of representations in general, and not just mental representations.

Numerous attempts have been made to provide explanations for why children acquire such an understanding of representations. Frye, Zelazo and colleagues (Frye, Zelazo & Burack, 1998; Frye, Zelazo & Palfai, 1995; Zelazo, Burack, Benedetto & Frye, 1996, Zelazo, Burack, Boseovski, Jacques & Frye, 2001; Zelazo & Frye, 1998, see also Perner & Lang, 2002 and Perner, Stummer & Lang, 1999 for a critique) have advocated an approach that views the development of a representational ‘theory of mind’ as a specific consequence of the operation of domain-general processes. They argue that once children understand that when ‘if-then’ rules are hierarchically embedded, successful resolution of, inter alia, unexpected transfer false belief tasks becomes possible. In such tasks, the statements forming the two rules are ‘[protagonist] knows where marble is’ and ‘I [child] know where marble is’. The difficulty, according to Frye and Zelazo is that younger children have difficulty hierarchising these statements into a structure whereby one acts as a setting condition for the other, a difficulty that is not restricted to situations involving mental state reasoning. Frye et al. (1995) report a study in which children had to predict the direction a marble would take down a covered ramp depending on whether a lamp was lit or not. Three-year-old children always predicted that the marble would follow the path dictated by gravity, whereas four-year-olds could modulate their predictions according the state of the light. Moreover, a strong correlation was found between performance on this task and performance on an unexpected transfer false belief task. This finding was replicated by Zelazo, Jacques, Burack and Frye (1996) using children with Down syndrome and, in a study of children with autism, Zelazo et al. (2002) also report strong correlations between false belief task performance and the ramp task, as well as with a dimensional change card sorting task, which requires children to sort cards according to colour or shape depending on the experimenter's instructions. Younger children with typical development and children with autism (both characteristically false belief failers) found it hard to shift sorting dimension in response to the experimenter's instructions. Although they could sort red and blue triangles and circles according to a pair of rules such as 'if this is red, put it in the red pile, if it is blue, put it in the blue pile', they found it hard to switch to putting triangles with triangles and circles with circles when asked to do so by the experimenter. Frye and Zelazo conclude from these studies that failure on false belief tasks is but one example of a more general executive failure to switch flexibly between rules. In their Cognitive Complexity and Control Theory, Zelazo and Frye (1998**,** Zelazo et al., 2001) argue that such flexibility is the result of increased psychological distancing from the problem that enables greater conscious control to be exercised over complex problems.

The theme of executive processes as contributing to performance on false belief tasks has also been explored by Russell (1996, 1998a,b). Russell starts from the observation that individuals with autism, who generally experience difficulty on unexpected transfer false belief tasks (Baron-Cohen, Leslie & Frith, 1985), are impaired compared to controls on some tasks of executive function but not on others. Tasks that show impairments include the Wisconsin Card Sorting Test, the Tower of Hanoi and its variants (Ozonoff & McEvoy, 1994), whilst unimpaired performance has been found on the stroop task (Bryson, 1983). Russell (1998b) argues that what distinguishes those executive tasks on which individuals with autism are successful from those on which they are not is the question of arbitrary rules; individuals with autism perform poorly on tasks that contain both arbitrary rules and require the inhibition of a prepotent response. So, they are impaired on a task where successful retrieval of a marble from a box has to be preceded by the arbitrary act of throwing a switch on the side of the box (Hughes & Russell,1993) or by turning over a cup placed beside the apparatus (Biró & Russell, 2001). The requirement to perform a non-arbitrary act such as removing an obstacle did not impair performance.

In applying this analysis to False Belief and False Photograph tasks, Russell concludes that the rules linking the elements of these two experimental scenarios do not bear an arbitrary relation to the events narrated, thereby lessening the executive demands of the tasks. But, Russell argues, although the two tasks contain aspects that might elicit prepotent responses (i.e. there is a current reality that conflicts with a representation of a past reality), this conflict is stronger in the case of false belief than false photograph tests. It is this weaker conflict between actual and represented reality that explains why children with autism are far more successful on tests of false non-mental representations such as out-of-date photographs (Leekam & Perner, 1991; Leslie & Thaiss, 1992) or out-of-date drawings (Charman & Baron-

Cohen, 1992), all of which have the same logic as Zaitchik's false photograph task described earlier.

The décalage between the false belief and false photograph performance of children with autism has formed the basis of an alternative, more domain-specific, modularist view of the development of an understanding of mind. Starting from Baron-Cohen et al.'s (1985) observation that mental state understanding either fails to develop or is markedly delayed in children with autism, Baron-Cohen (see Baron-Cohen, 1995) concludes that children with autism lack a 'theory of mind', that is, that they are specifically impaired in their understanding of mental states. Baron-Cohen (1995) further argues that this impairment was specific to mental states because even those children with autism who fail false belief tasks have little difficulty with tests of non-mental representation (see above).

The mental state specificity argument proposed by Baron-Cohen is based on two assumptions. The first is that unexpected transfer false belief tasks and Zaitchik's (1990) out-of-date or 'false' photograph task are equivalent in all respects except for their mental state content. The second is that the unexpected transfer false belief task is an unambiguous measure of mental state understanding. However, a consideration of the structures of the two tasks calls these assumptions into question. Whereas false photograph tasks involve the experimenter, an object, two locations and a representation of the object, false belief tasks typically involve the experimenter, two protagonists, an object and two locations. Moreover, the child has to take into account the behaviour of the protagonists towards the object. The false photograph task involves fewer elements and actions than the false belief task and so can be regarded as less complex and therefore less demanding on the child. Furthermore, whilst unexpected transfer tasks can undoubtedly be thought of as measures of mental state understanding, they can also be conceptualised in a way that makes no reference whatsoever to mental states. They can be viewed in terms of an agent's goal-directed behaviour that is mediated not by the state of the goal, but by a device that can truly or falsely signal the state of the goal to the agent (see Table 1). In the typical Sally-Anne version of the unexpected transfer scenario (Baron-Cohen et al., 1985), we have an Agent (Sally) who behaves towards a particular Goal (marble), which can be in one of two locations (in the box or in the basket). Each of these two locations can be registered by a Signal (Sally's belief), which determines the Agent's (Sally's) behaviour towards the Goal (Marble) and which, in turn, is determined by the location of the Goal. The state of Sally's belief (the Signal) is usually, but not necessarily congruent with the location of the Goal. So, when the Signal and the location of the Goal are in conflict (i.e. when Sally believes something that is not actually true), then we would predict that a signal-dependent agent (Sally) should obey the Signal (her belief) and not the location of the Goal. So a child who passes such a test has to grasp three things: that agents act towards goals, that these acts can be mediated by signals and that such signals can be false.

If the above analysis of false belief tasks is true, then it would follow that any task that was consistent with that analysis should correlate highly with performance on the Sally-Anne false belief task. Moreover, any such correlation would weaken the argument that passing false belief tasks involves specifically an understanding of mental states. Such a correlation would implicate an understanding of the behaviour of agents towards objects, which Leslie (1987, Leslie & Roth, 1993) have argued is necessary for the understanding of other minds. It would also imply an understanding that behaviour in relation to a goal can be mediated by a signal that truly or falsely represents the state of that goal. But it would not limit such understanding specifically to mental states. On the other hand, the implications of failing to find a correlation would depend on the precise pattern of results obtained. A random pattern of responses would indicate that children failed to understand the task and higher or lower performance on the analogue task would suggest that it was assessing different capacities from those assessed by the false belief tasks. A similar set of arguments holds for the case of autism. In particular, a significant correlation between the two tasks would suggest that the delay or failure to master false belief tasks seen in this population may not be due to a specific deficit in mental state understanding but rather to a more general difficulty with systems involving signal-dependent agents.

This analysis of unexpected transfer tasks in terms of Agents, Signals and Goals may also help to explain why false belief and false photograph tasks are correlated in typically developing children yet not correlated in children with autism. Both tasks entail a grasp of a signal truly or falsely representing a goal, but only the false belief task entails an agent acting on the basis of this information. It may be the case that typically developing children understand both these aspects of the problem, whereas children with autism understand only the first – i.e. they fail to understand agents' goal-directed behaviour. As a consequence, they perform well on tasks involving a false signal (e.g. false photograph tasks) but poorly on those that require the prediction of agents' behaviour (e.g. false belief tasks).

Experiment 1

To test the possibility that failure on false belief tasks may simply be the result of a difficulty in prediction of an agent's goal-directed behaviour on the basis of incorrect information signalling the state of the goal, we developed a mechanical analogue of the standard, unexpected transfer false belief task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983). In this task, children had to predict the behaviour of a mechanical system that used signal information about an event on which that system had to act (see right-hand column of Table 1). We predicted that if performance on the standard, unexpected transfer false belief task depends on an ability to predict an agent's behaviour on the basis of signal information, then there should be a strong correlation between the analogue task and the standard false belief task. Failure to find a correlation between the two tasks would be consistent with the idea that there is a specific mental state component to false belief tasks. We also predicted that if children with autism have a specific deficit in understanding mental states, then their performance on the mechanical analogue would be uncorrelated with false belief task performance.

Method

Participants

There were three groups of participants: children with autism, children with moderate intellectual disabilities and typically developing children. The children with intellectual disabilities were included to control for the fact that children with autism frequently are of lower IQ than typically developing children. The children with autism all had a diagnosis of autism according to DSM III-R criteria ascertained from school records and from direct observations by the first author. Children with intellectual disabilities were excluded if there was any mention of autism or autistic-like behaviour on the school records of if there was any present-state evidence of autistic social impairment. The typically developing children and those with intellectual disabilities were all selected from nursery and special schools in London. All but 13 of the children with autism were selected from special schools in London and the south-east of England. For all these children, verbal mental ages were measured using the British Picture Vocabulary Scale (BPVS, Dunn, Dunn, Whetton & Pintillie, 1982). The remaining children with autism, five of whom had French as a first language were selected from a school in Montreal. Verbal mental ages for these children were measured using a French-language version of the Peabody Picture Vocabulary Test and appropriate test procedures and norms. Details of chronological and verbal mental ages are set out in Table 2. One-way ANOVAs revealed a significant difference for Chronological age only (F (2, 64) = 76.74, $p < 0.001$). Post-hoc Scheffé tests revealed significant pair-wise differences in chronological age for all three pairs of groups.

Because the present study was designed to investigate correlations between two tasks, it was important to avoid both ceiling and floor effects in performance on the two measures. To avoid such effects, participants from all three groups were selected so as to have a reasonable spread of passers and failers on the Sally-Anne task consistent with good matching on verbal mental age. This selection was done by means of a process of progressive inclusion, whereby the group pass and fail rates on the Sally-Anne task were evaluated from time to time. When there appeared to be a preponderance of passers on this task (> 75%), additional children with lower VMAs were recruited. When failers were in a majority, children with higher VMAs were selected. All this selection was carried out without regard to performance either on the Sally-Anne or theTrain task, and resulted in the verbal mental age of the typically developing group being a marginally lower and that of the autism group marginally higher than that of the other two groups.

Equipment

The Sally Anne false belief task was administered using two small dolls and a box, basket and marble. The mechanical analogue (Train task) used the equipment pictured in Figure 1. The apparatus, which measured 108 by 60 cm was built using 0-gauge model railway track and consisted of a model airport with two landing pads, one coloured yellow and one blue, as well as one passenger terminal. Each landing pad was surrounded by a nine cm high wall painted in the same colour as the pad, and could be covered by a lid of the same colour. A fully automatic, driverless shuttle conveyed goods from terminal to aircraft. At first the shuttle followed a single path from the terminal but some way from the terminal the track divided and the route to be taken was signalled by a yellow or blue light activated by the plane landing on the yellow or blue pad respectively. A console under the experimenter's control regulated train movements and points and signal operations.

Procedure

Train Task

Initially, children were tested using the scenario pictured in Figure 1 to ensure that they had understood that a plane landing on yellow caused the yellow light to show and that a yellow light showing meant that the shuttle headed for the yellow pad etc. It was emphasised to the children that the train was fully automatic, had no driver and relied on the signal to determine which direction it should take.

Once children had demonstrated that they reliably understood these features of the system by making five correct predictions out of six trials, the experimenter (unknown to the child) switched the mechanism relating the yellow and blue lights to the landing pads so that when a plane now landed on the yellow pad, a blue signal light came on and vice versa. This change was accompanied by expressions of surprise by the experimenter who pointed out the contrast between the colour of the location of the plane and that shown by the light. At this point, the experimenter covered the landing pads with their lids and switched off the signal light. Children were then asked to predict where the shuttle would go, given that the yellow light was showing and that they could see that a plane had landed on the blue pad. Once they had answered the prediction question, children were also asked about the location of the plane and the colour of the light. One trial was given, with the colour of the pad counterbalanced across children. The text of the instructions given to the children can be found in the Appendix.

Sally-Anne Task

All children were also given one trial on a version of the Sally-Anne unexpected transfer, false belief task used by Baron-Cohen et al. (1985) to assess understanding of false belief, with the location of the marble counterbalanced across children. The test question used here was 'Where will Sally look for her marble?' and was followed by reality and memory control questions in counterbalanced order. Order of administration of Sally-Anne and Train tasks was counterbalanced across children. The instructions given to the children are in the Appendix.

Results

On both tasks, children were included only if they correctly answered the Reality and Memory control questions on the Sally Anne task and the Plane Location and Colour of Light question on the Train task. Four typically developing children, six with intellectual disabilities and five children with autism were excluded because they failed one or more control questions. The Sally-Anne task was scored by assigning 0 for a fail and 1 for a pass. For the Train task, it was decided to assign 0 if a child made a 'plane' response and 1 if a 'light' response was made. 'Plane' and 'light' are used here rather than pass/fail because it is less clear in this task that one or other option could be considered as genuinely successful. The results of the data analyses are the same whichever way the data are coded.

The data were tested first for sampling and order effects, followed by between-group, within-task comparisons using chi-square tests. A series of within-group, between task tests were performed in order to explore associations between the tasks using the Phi test, as well as to compare the difficulty of the two tasks using the McNemar test. Finally, to examine the relative contributions of the Train task and verbal mental age to Sally-Anne performance, we carried out a logistic regression analysis.

To test for the effects of order on task performance, a series of chi-square tests was carried out on the numbers of passers and failers on each task in each counterbalancing group separately for each group of participants. All comparisons were non-significant ($\chi^2 < 1.05$). Comparison of the UK and Montreal children with autism on CA, VMA and the two experimental tests revealed no significant differences between these two groups justifying their inclusion in a single sample. A set of pair-wise, between-group comparisons were carried out on the Sally-Anne and Train pass and fail rates that are set out in Table 3. In all cases $\chi^2 < 1$, indicating similar patterns of performance in the three groups and the proportion of Sally-Anne task failers in the autism group, at 33%, is in line with earlier studies (see Happé, 1995).

Set out in Table 4 are the within-group cross-tabulations of pass and fail rates on the Sally-Anne and Train tasks. Inspection of these data show that both tasks are strongly correlated in all three groups of children (all Φ-values >0.55) and that the two tests are also of equal difficulty (all McNemar tests ns). Both these observations strongly support the rationale behind the development of the Train task and the contention that this task is a good analogue of the Sally-Anne false belief task for all three groups of participants.

In view of the strong relation between verbal mental age and performance on false belief tasks reported by Happé (1995), a series of correlations between performance on the two experimental tasks and BPVS age-equivalents were calculated. A significant correlation was found between Verbal Mental Age and performance on the Sally-Anne task ($r = 0.28$, N = 67, $p < 0.05$), but not for the Train task ($r = 0.22$, $N = 67$, ns) . In view of this correlation, a logistic regression analysis was carried out to test the extent to which scores on the Train task predicted Sally-Anne scores for all the children used here after the effects of verbal mental age had been removed. As verbal mental age has been demonstrated by previous research to correlate with false belief task performance (Happé, 1995), we followed Field (2000) and entered this variable on the first step of the regression analysis. Inclusion of BPVS age-equivalent significantly improved the predictive power of the constant-only model (χ^2 (1, $N = 67$) = 5.20, $p < 0.03$) yielding successful classification rate of 60%. Addition of the Train task data in the second step of the model further improved the power of the model (χ^2 (1, $N = 67$) = 28.92, $p < 0.01$) and increased the successful classification rate to 85%. Regression statistics for BPVS age-equivalent were $\beta = 0.04$; Wald z-ratio = 2.3, ns; Odds Ratio = 1.04. The corresponding values for the Train task score were $\beta = 3.20$; Wald z-ratio = 21.61, $p < 0.01$; Odds Ratio = 24.44. Inspection of the Wald criterion scores shows that performance on the Sally-Anne task is reliably predicted only by the Train task data.

Diagnostic statistics from the regression analysis revealed a good fit between the model and the data. Five children (two with typical development, two with intellectual disabilities and one with autism) who passed the Sally-Anne task were mis-classified as failers by the regression model and a further five children (one with typical development, one with intellectual disabilities and three with autism) who failed the task were mis-classified as passers. The remaining 57 children were correctly classified. All Cook's distances were <1.00, a value which Howell (1997, p.535) suggests as a cut-off value for outliers. Only one leverage value was greater than twice the average value, a level normally taken to be a measure of outliers (Hoaglin & Welsch, 1978 cited in Field, 2000). In addition, none of the DF-beta values for the constant or the two predictor models exceeded 1.1.

Discussion of Experiment 1

The finding of a significant association between the two tasks for all three groups of participants supports the initial rationale behind the development of the Train task and suggests that the two measures tap some underlying capacity to understand how the behaviour of an agent towards one event can be influenced by another event that can stand in true or false relation to the first. The absence of any mention of belief, the absence of people from the Train scenario make it unlikely that the cognitive capacity common to both tasks entails an understanding of mental states.

The lack of significant differences in Table 3 between either the typically-developing children or those with intellectual disabilities and the children with autism appears to contradict existing research in which impairment in the last group is the rule (e.g. Baron-Cohen et al., 1985). However, the success rate of the children with autism (33%), although high by comparison with some studies, does lie within the range reported in other studies (see Happé, 1995). These pass rates on the Sally-Anne task that are lower than usual for the typically-developing children and higher than usual for the children with autism are partly due to the sampling strategy described earlier on, which was designed to achieve a reasonable spread of Sally-Anne task passes and failers in all three groups.

Although the findings of the present experiment are consistent with the view that failure to understand false beliefs in individuals with autism may not be the result of an impairment of a mechanism specific to the understanding of mental states, there is one respect in which the Train task is not an exact analogue of the Sally Anne task. In the latter task, the event that engenders Sally's false belief is Anne's transfer of the marble. In the Train task, there is no event that is an exact counterpart of this action. Failure to include this element of the task may have made the Train task easier than it would otherwise be, and so have contributed to artefactual correlations between the two tasks. To overcome this possible source of error, we developed a modification of the Train task for use in Experiment 2.

Experiment 2

The points raised in the discussion of Experiment 1 prompted a re-design of the procedure used in Experiment 1. The aim was to include an explicit event corresponding to Anne's transfer of the marble in the Sally-Anne test that sets up the conflict between the location of the plane and the state of the signal It is possible that the omission of this element from the Train task made it easier than the Sally-Anne task thus making the correlations found in Experiment an artefact of the procedure rather than a genuine finding.

Method

Participants

Eighteen children with autism and 25 typically developing children were selected from nursery and special schools in the Greater London area. Selection criteria were the same as those used in Experiment 1 and no child took part in both experiments. Because there were no differences between the children with autism and the children with intellectual disabilities in Experiment 1, a group of the latter children was not recruited for this experiment. Details of chronological ages and verbal mental age equivalents, measured by the BPVS are set out in Table 5. Separate independent-sample t-tests revealed no significant difference between the two groups on VMA ($t < 1$) but a significant difference on CA ($t (41) = 8.98$, $p < 0.001$).

Equipment

The Sally-Anne false belief task was administered using two small dolls and a box, basket and marble, exactly as in Experiment 1. The Train task used the same equipment as used in Experiment 1 with the addition of a small toy bird.

Procedure

Train Task

The initial familiarisation procedure and criterion used here were identical to those in Experiment 1. However, on the test trial, as the plane was about to land on the yellow (or blue) pad, a small bird landed there first, thereby activating the yellow (or blue) light. The experimenter then pointed out that the bird was stopping the plane from landing on that pad and that it would have to go and land on the other one. The plane was then made to land on the other pad. The signal light continued to show the colour of the first pad. The experimenter then pointed out that the plane was on one colour while the signal showed the other. Each child was given one trial, with choice of landing pad counterbalanced across children. Children were then asked, in counterbalanced order, to indicate the actual location of the plane and the colour that the light had been showing.

Sally-Anne Task

All children were also given one trial on the same Sally-Anne task used in Experiment 1. The order of presentation of the Train and Sally-Anne tasks was counterbalanced across children.

Results

The same scoring procedures and data analytic strategy was used as in Experiment 1. Unfortunately, systematic records were not kept of the numbers of children excluded because of incorrect answers to control questions, so these cannot be reported here. To test for the effects of order of presentation on task performance, a series of four chi-square tests were carried out between order of presentation and score on each of the two tests, separately for the two groups of participants. In all cases $\chi^2 < 1$ indicating that the order of testing did not affect performance on either test.

Pair-wise between-group comparisons of pass and fail rates on the Sally-Anne and Train tasks set out in Table 6 indicate similar patterns of performance in both groups (in all cases $\chi^2 < 1$). The proportion of Sally-Anne task passers in the autism group, at 50%, is higher than that reported in other studies, but must be taken in the context of the sampling issues mentioned in the discussion of Experiment 1.

Set out in Table 7 are the within-group cross-tabulations of the Sally-Anne and Train tasks. Inspection of these data show that both tasks are strongly correlated in both groups (both Φ-values >0.58) and that the two tests are also of equal difficulty (both McNemar tests ns). These observations replicate the findings of Experiment 1 and strongly support the rationale behind the development of the Train task and the contention that this task is a good analogue of the Sally-Anne false belief task for both groups of participants.

For the same reasons as in Experiment 1, correlations were calculated between performance on the two experimental tasks and BPVS age-equivalents. Both values were significant (Sally-Anne task: $r = 0.32$, $N = 43$, $p < 0.05$; Train task, $r = 0.41$, $N = 43$, $p < 0.01$). A logistic regression analysis was then carried out using the same procedure as in Experiment 1. Inclusion of BPVS age-equivalent on the first step significantly improved the predictive power of the constant-only model (χ^2 (1, $N = 43$) = 4.85, $p < 0.03$) yielding successful classification rate of 65%. Addition of the Train task data in Step 2 further improved the power of the model (χ^2 (1, $N = 43$) = 15.62, $p < 0.01$) and increased the successful classification rate to 84%. Regression statistics for BPVS age-equivalent were $\beta = 0.14$; Wald z-ratio = 0.16, ns; Odds Ratio = 1.01. Corresponding values for the Train task were $\beta = 3.10$; Wald z-score = 11.92, $p < 0.001$; Odds Ratio = 22.13. In replication of Experiment 1, the Wald criterion scores indicate that performance on the Sally-Anne task is reliably predicted only by the Train task data.

As in Experiment 1, diagnostic statistics from the regression analysis showed that the model fitted the data well. Four children (two from each group) who failed the Sally-Anne task were predicted as having passed the task and three children (all with typical development) who passed were predicted as having failed. Predictions for the remaining 36 children were in line with their actual test performance. All Cook's distances were <1.00, three leverage values were greater than twice the average value and none of the DF-beta values for the constant or the two predictor models exceeded 1.5.

Discussion of Experiment 2

By replicating the findings of Experiment 1, the results of the second experiment reinforce the validity of the rationale behind the construction of the Train task. The addition of a step in the Train task corresponding to Anne's displacement of the marble in the Sally-Anne task did not affect the overall thrust of the results. This finding further reinforces the conclusion that what is important in this task is the mis-match between the location of the plane and the colour of the light indicating that location. The fact that in both experiments, the Sally-Anne false belief task correlated highly with an entirely mechanical task, irrespective of the diagnosis of the children, suggests that failure on the former task may not be due to a specific failure to understand mental states but rather some more general difficulty in understanding signal-mediated behaviour. Although the correlational data presented here do not permit the inference of direction of cause between tasks, (see Colvert, Custance & Swettenham, 2002 for a discussion of correlation issues in relation to false belief tasks) there are aspects of the two tasks and of the data that allow some speculations to be made. These will be discussed in the next section.

General Discussion

The main findings of the two experiments reported here are that it is possible to devise experimental scenarios on which children perform in a way that correlates highly with their performance on false belief tasks despite the former having no explicit mental state content. There are three possible explanations of the correlations reported here between the Sally-Anne and Train tasks. Either the first causes the second, the second causes the first or they are both caused by a third factor or set of factors. To make the first argument, that is to assert that an understanding of mental states, or the possession of a ‘theory of mind’ is necessary in order to pass the Train task is to come close to the ideas of Perner (1991), who would argue that the signal in the Train task is a representation of the plane’s location, and that only children who understand the representational nature of the signal (i.e. that it can mis-represent the location of the plane) will pass the task. However, such an argument reduces the problem in both tasks to one of representation in general, rather than one specifically of mental states. But children with autism typically perform much better on tests of understanding of non-mental representation such as false photographs or maps (Charman & Baron-Cohen, 1992;

Leekam & Perner, 1991; Leslie & Thaiss, 1992) than they do on false belief tasks. By extension, if the problem were simply one of understanding non-mental representations, children's performance on the Train task used here should have been better than on the Sally-Anne task.

In defence of findings that show superior performance by children with autism on false photograph tasks over false belief tasks, Perner (1993) has argued that such children are generally older than the typically developing controls and thereby have had more experience of cameras and photographs. In the present context, such an argument would explain the equivalent performance of the children with autism and typically developing children on the Train and Sally-Anne tasks in terms of equivalent experience with signal-dependent trains. But apart from having an unsatisfactory post-hoc feel to it, such an argument also seems to ignore the fact that all children (whether with or without autism) are likely to have had far more experience of people than cameras or toy trains and thus should perform better on tests of false belief than on tests of artifactual false representation such as false photographs or drawings. The fact that they do not suggests that an analysis in terms of representation may not be the most appropriate to explain either the present findings or those from false photograph studies.

It can also be argued that the correlations between the two tasks reflect their common mental state content on the basis of a Thomas the Tank Engine factor[1]. Throughout, we have argued that the Train task does not require any mental state

---

[1] We are grateful to Greg Currie for bringing this objection to our attention.

understanding for its solution. But children may well either assume that the train has a driver, or that it has a personality of its own, thus making it a test of understanding mental states and rendering any correlation with the Sally-Anne task unsurprising and theoretically uninteresting. Although it is true that trains and their drivers figure strongly in the cultural environment of children and that many of the participants used here may well have known of Thomas the Tank Engine, it is unlikely that these factors influenced the findings of the present study. First, it had been strongly emphasised to children that the train had no driver to see where the plane had landed and that a signal was needed to indicate which direction to take. Second, the training procedure, in which the experimenter made the train travel to and from the actual location of the plane using a control box with switches and knobs, reinforced the mechanical nature of the set-up. This procedure made the scenario resemble a 'Scalextric' car-racing toy rather than something inhabited by person-like entities. Third, the train engine used was a model of a modern electric locomotive, devoid of the characteristic features (e.g. a highly differentiated front and back, round features on the front that might suggest a face) that would prompt the attribution of personality. But the most important evidence against attribution of personality to the train lies in the data. If children had assumed either that the train had a driver or that it had its own personality, they would never have predicted that the train would have followed the signal, because the location of the plane was clearly visible from the starting point of the train. It can be objected that many children did just that, namely that they predicted that the train driver would ignore the signal and head for the actual location of the plane. But such an objection needs to explain why it was only the Sally-Anne failers who did this. Sally-Anne passers did not. To account for this pattern in the data, a proponent of the Thomas the Tank Engine argument would need to explain how the acquisition of an understanding of false belief results in children no-longer attributing personality to the train - a sort of reverse 'theory of mind'. Further evidence on this point could be gained from asking children to justify their choice of responses on the two tasks to see whether they made mental state attributions on the Train task. It might also be useful to devise a scenario using animated coloured shapes or blocks that had a similar logical structure to that of the train task. Whether or not children would make sense of such abstract scenarios is an open question.

Related to the Thomas the Tank Engine factor is a more general argument that both tasks used here are simply alternative measures of ‘Theory of Mind’ and that children confronted with either of them will respond according to their level of development in that domain. Such an argument leads to the prediction that the two tasks would be strongly correlated and of equal difficulty – precisely the findings reported here. But such an argument seems to stretch the domain of ‘mind’ beyond what might reasonably be thought of as the realm of mental phenomena. If we were to accept this argument, then we would have to accept that signal-dependent systems of goal-directed action are ‘mental’ systems. The conjecture made in this paper is that ‘mental’ systems are a subset of signal-dependent goal-directed systems and that when children understand one system, they should automatically understand the other.

If, however, we restrict our conception of mind to unobservable, hypothetical entities in the heads of persons, then there are two other ways in which possession of a 'theory of mind' could be thought of as the causal factor in passing the Train task. If we assume that 'theory of mind' capacity once developed then gets applied, by analogy or some other process, to situations not involving mental states, we would predict a number of specific patterns in the data. Because performance on the Train task would depend on the application of 'theory of mind', i.e. would be developmentally secondary to theory of mind, we should expect performance on the Train task to be worse than that on the Sally-Anne task. However, inspection of the data in Tables 4 and 7 shows that this is not the case for any group in either experiment. Moreover, if successful performance on the Train task required an intact 'theory of mind', then responses on this task by Sally-Anne task failers should be distributed randomly between the plane and the light and, additionally, not all Sally-Anne passers should choose the light, because some passers would not yet have learned to apply their mentalising skills to non-mental situations. Inspection of Tables 4 and 7 supports neither of these possibilities. A significant proportion of Sally-Anne failers in all these groups chose the plane rather than the light, and a significant proportion of Sally-Anne passers (over 80% in each case) chose the light. The corresponding figure for the children with autism from Experiment 1 (75%) was not significant, but was nevertheless a majority and may have been affected by sampling considerations that yielded a relatively low number of false belief passers in this group. Once a sample of children with autism with a higher false belief pass rate was recruited, as in Experiment 2, a similar effect to the other groups was found.

To make the reverse of the argument just presented, namely to assert that an understanding of signal-dependent behaviour of agents causes an understanding of false belief would make the opposite predictions from those just outlined. So, we would expect the responses of children who chose the plane to pass or fail the Sally-Anne task in equal numbers, and for a substantial number of those who chose the light to continue to fail the Sally-Anne task. Inspection of Tables 4 and 7 reveals none of these eventualities in the data.

A further possibility to account for the correlation between the two tasks is that they both depend on some common factor. In view of Happé's (1995) meta-analysis of factors that influence performance on false belief tasks, the most obvious contender in this respect is verbal mental age. However, although the correlational analyses show that verbal mental age correlates with False Belief task performance in Experiment 1 and with both tasks in Experiment 2, the logistic regressions show that the two experimental tasks are still correlated with each other when the effects of verbal mental age were partialled out. Another possible common factor is the ability to disengage from salient reality in the context of an arbitrary rule (Russell, 1996). In the present two scenarios, someone looking for a marble will go to where the marble is, and a train that picks up goods from a plane will travel towards the plane. This account would suggest that the Train task requires participants to disengage from a salient reality (the current location of the plane) in the context of a rule (if the light is yellow, then go to the yellow pad...), thereby making the task as difficult as the false belief task. If this account is true, then manipulations of the salience of the plane or the signal as well as the changing arbitrariness of the rule linking the signal to the location of the plane should alter success rates on this task.

According to the analysis of the Train and Sally-Anne tasks in terms of signal-dependent, goal-directed behaviour presented earlier, it is plausible to argue that the factor common to the two tasks is an ability to manipulate information about the state of the goal, the state of the agent and the mediating signal that conveys information to the agent about the state of the goal. Although we have not conclusively established that an agent is a necessary component of this system, we will use that term for any entity that can change state systematically in response to state changes in another entity. The crucial factor here is that we have three binary-state elements, with the state of the first (the goal) being mirrored by the second (the signal) but with the possibility of such mirroring being wrong, and a third which changes in response to changes in the first on the basis of information provided by the second. On this analysis, children's difficulty with false belief is not specific to understanding mental states, but rather, is a more general information-processing deficit involving the manipulation of contingent information from more than two sources in a way that allows the state of a third element of the problem to be predicted.

The work of Frye, Zelazo and colleagues (Frye et al., 1988; 1995; Zelazo & Frye, 1998; Zelazo et al., 1996; 2001; 2002) on embedded conditional reasoning described earlier is consistent with the analysis just presented. In their dimensional change card sorting test the child has to consider in which of two locations to place an object based on two other binary sources of information (shape versus colour game and either circle/triangle or red/blue). It is possible to conceptualise the Train task used here in embedded conditional terms. A single pair of rules could be either 'If the plane is on blue, then the train goes to blue; if the plane is on yellow, the train goes to yellow'. A similar pair of rules operates for the yellow and blue lights. The setting rule, analogous to Frye and Zelazo's colour/shape game instruction, would be 'If the colour of the light and the colour of the landing pad where the plane is are the same, then use the first pair of rules; if they are different, use the second. If this analysis were correct and conditional rule use is a common factor underlying performance on Frye and Zelazo's tasks, false belief tasks and the Train task used here, then we would predict strong inter-correlations among these tasks for all groups of participants used here.

Another account of false belief understanding in terms of the manipulation of complex information has appeared in the literature in recent years, which has aspects similar to the analysis just outlined. Halford (1993, 1996, Halford, Wilson & Phillips, 1998) argues that all problems can be analysed in terms of their processing complexity, defined as '... the number of interacting variables that must be represented in parallel to perform the most complex process involved in the task...' (Halford et al., 1998, p. 805). Halford and his colleagues argue that cognitive development involves a progression from an ability to process single elements (unary) through two-element relations (binary) to three-element relations (ternary) and beyond. Successful resolution of false belief tasks involves an understanding of ternary relations by the successful co-ordination of pairs of binary relations. According to this view, younger children (and most children with autism) have little difficulty in understanding any binary relationship between Sally and the location of her marble. That is to say that they can grasp either that ‘Sally thinks her marble is in the box’ or ‘Sally thinks her marble is in the basket’. They experience difficulty, however, when they have to decide between these two relationships on the basis of an event that transforms one into the other – in this case, whether or not Sally saw the marble being moved. Although Halford provides numerous empirical examples from outside the area of mental state understanding in support of his theory, so far, none has been developed in the context of false belief understanding. The studies described here, although conceived in ignorance of Halford's theory, provide evidence that is consistent with it.

Whichever of the accounts just outlined turns out to be true, the further question of what it is that is different about the children with autism who, even when they pass higher-level tests of ‘theory of mind’, remain autistically socially impaired (Bowler, 1992; 1997; Ziatas, Durkin & Pratt, 1998). Related to this is the question of why children without autism who can not process information in the way described in the previous paragraphs are not autistic. One possibility that has been raised by a number of authors in the context (Baron-Cohen, 1989; Happé, 1994) of the first point is that individuals with autism are delayed in their acquisition of these abilities and that they never ‘catch up’ to a normal level of social functioning. However, empirical support for such a conjecture is mixed. The results of the present study lend some tentative support. Inspection of Tables 2 and 5 shows that the mean verbal mental age of the children with autism is 10 months (Experiment 1) and six months (Experiment 2) higher than that of the controls, which confirms the finding by Happé (1995) of a higher mean VMA needed to pass false belief tasks by such children .

The fact that typically-developing children who fail the Sally-Anne task are not autistically socially impaired suggests that this task, and by implication, the three-element schema outlined earlier on is only part of the explanation for normal social functioning, and prompts a consideration of other possible elements in the process. Leslie's (1987; Leslie & Roth, 1993) two-step account of the development of false belief understanding may be of relevance here. Leslie has argued that the capacity to generate agent-centred and possibly counter-factual representations develops in infancy and enables pretence, whereas the mechanism that enables the prediction of behaviour on the basis of false belief (the SP mechanism) does not emerge until about four years of age. It is possible that the three-element schema outlined here is a candidate for the internal workings of the SP. If this is so, then what is lacking in children with autism is, as Leslie and Roth (1993) and Leslie and Thaiss (1992) point out, an understanding of agent-centred representations. Two strands of evidence lend support to this conjecture. The first is from Tables 3 and 7, which show similar levels of performance by children with autism on the mental-state, “agent-ful” Sally-Anne task and the non-mental-state, “agent-ful” Train task. It will be remembered that children with autism typically do better on “agent-less” non-mental-state tasks (false photographs or maps) than on false belief tasks (Leekam & Perner, 1991; Leslie & Thaiss, 1992). So the inclusion of an agent brings the performance of children with autism on non-mental-state tasks back to the level of their performance on false belief tasks. This argument would make the prediction that children with autism would perform better on an “agent-less” system that operated on the three-element schema than they would on Sally-Anne type false belief tasks.

A second strand of evidence that suggests that children with autism have problems understanding agents comes from a study by Bowler and Thommen (2000), who asked such children together with controls matched on VMA, CA and IQ to describe a short animated film developed by Heider and Simmel (1944). In the film, three animated (self-propelled) geometric shapes interact with each other and with a static rectangle. Analysis of children’s descriptions of the film revealed that the children with autism were delayed in their ability to describe relations between an animate character and the inanimate object. Moreover, they were worse than all the control groups at describing actions between two animate objects when they did not come into contact (e.g. a chase). These findings suggest that children with autism fail to notice certain events in their environments that help to build up an understanding of the relations among agents and between objects and agents and thus have difficulties on all tasks including the two used here, where agents’ goal-directed actions have to be predicted. Such a failure to understand the actions of agents at a distance is in line with the developmental symptomatology of autism (e.g. absence of protodeclarative pointing, see Mundy, Sigman & Kasari, 1993). Moreover, the strong contingency between the two tasks in the two experiments reported here supports their having a common element such as agency. Like the argument developed in the previous paragraph, this account makes the prediction that the performance of children on “agent-less” tasks would be better than on “agent-ful” ones. The suggestion the problem faced by children with autism on false belief-type tasks resides in difficulty with the comprehension of the behaviour of agents also helps to resolve the discrepant correlations observed in this group and in typically developing children between false belief and false photograph tasks. Whereas ability to succeed on both tasks comes on line at the same time for typically developing children, children with autism can pass false photograph (“agent-less”) before they can pass false belief (“agent-ful”) tasks.

To conclude, the findings of the present investigation suggest that the failure of children with autism on unexpected transfer false belief tasks is unlikely to be due to a specific deficit in the understanding of mental states but rather, that such understanding may be a sub-set of comprehending specific types of complex situation. Although the present findings do not provide an unambiguous specification of such situations, they do narrow the set of possibilities down to a grasp of complex reasoning or of agency, or to a combination of the two. Complex reasoning may involve Halford’s (1993) notion of ternary relations, or Zelazo and Frye’s (1998) concept of hierarchically embedded rules. Alternatively, it may involve the executive processes of simultaneous disengagement from a salient object coupled with the operation of an arbitrary rule, as Russell (1998a) advocates. Further studies are needed to decide among these possibilities.

References

Baron-Cohen, S. (1995). Mindblindness Cambridge, MA.: MIT Press.

Baron-Cohen, S., Leslie, A. and Frith, U. (1985). Does the autistic child have a 'theory of mind'? Cognition, 21, 37-46.

Biró, S., & Russell, J.(2001). The execution of arbitrary procedures by children with autism. Development and Psychopathology, 13 , 97-110

Bowler, D.M. (1992). 'Theory of Mind' in Asperger's syndrome. Journal of Child Psychology and Psychiatry, 33, 877-893.

Bowler, D.M. (1997). Reaction times for mental-state and non mental-state questions in 'theory of mind tasks': Evidence against logico-affective states in Asperger's syndrome. European Child and Adolescent Psychiatry, 6, 160-165

Bowler, D.M. & Thommen, E. (2000). Attribution of mechanical and social causality to animated displays by children with autism. Autism, 4, 147-171.

Bryson, S.E. (1983). Interference effects in autistic children. Journal of Abnormal Psychology, 92, 250-254.

Charman, T. & Baron-Cohen, S.(1992). Understanding drawings and beliefs: a further test of the metarepresentation theory of autism. Journal of Child Psychology and Psychiatry, 33,1105-1112.

Colvert, E., Custance, D. & Swettenham, J. (2002). Rule-based reasoning and theory of mind in autism: A comment on the work of Zelazo, Jacques, Burack and Frye. Infant and Child Devlopment, 11, 197-200

Dunne, L.M., Dunn, L.M., Whetton, C. & Pintillie, D. (1982). British Picture Vocabulary Scale. London: NFER-Nelson.

Field, A. (2000). Discovering statistics using SPSS for Windows. London: Sage Publications.

Frye, D., Zelazo, P.D. & Burack, J.A. (1998). Cognitive complexity and control: I. theory of mind in typical and atypical development. Current Directions in Psychological Science, 7, 116-121.

Frye, D., Zelazo, P.D. & Palfai, T. (1995). Theory of mind and rule-based reasoning. Cognitive Development, 10, 483-547.

Halford, G.S. (1993). Children's Understanding: The Development of Mental Models, Hove, U.K.: Erlbaum.

Halford, G.S. (1996). Children's understanding of the mind: An instance of general principle? Contemporary Psychology, 49, 229-230.

Halford, G.S., Wilson, W.S. & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. Behavioral and Brain Sciences, 21, 803-864.

Happé, F.G.E. (1995). The role of age and verbal ability in the Theory of Mind task: performance of subjects with autism. Child Development, 66, 843-855.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. American Journal of Psychology, 57, 243-249.

Hoaglin, D. & Welsch, R. (1978). The hat matrix in regression and ANOVA. American Statistician, 32, 17-22.

Howell, D.C. (1997). Statistical Methods for Psychology (4th. ed.). Belmont, CA: Duxbury Press.

Hughes, C.H. & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. Developmental Psychology, 29, 498-510.

Leekam, S. & Perner, J. (1991). Does the autistic child have a metarepresentational deficit?. Cognition, 40, 203-318.

Leslie, A.M. (1987). Pretence and representation: The origins of "theory of mind". Psychological Review, 94, 412-426

Leslie, A. & Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg & D. Cohen (Eds.). Understanding Other Minds: Perspectives from Autism (pp. 83-111). Oxford: Oxford Medical Publications.

Leslie, A.M. & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. Cognition, 43, 225-251.

Mundy, P., Sigman, M. & Kasari, C. (1993). Theory of mind and joint-attention deficits in autism. In S. Baron-Cohen, H. Tager-Flusberg & D. Cohen (Eds) Understanding Other Minds: Perspectives from Autism pp 180-203). Oxford: Oxford Medical Publications.

Ozonoff, S. & McEvoy, R.E. (1994). A longitudinal study of executive function and theory of mind development in autism. Development and Psychopathology, 6, 415-431.

Perner, J. (1991). Understanding the Representational Mind. London: MIT Press.

Perner, J. (1993). The theory of mind deficit in autism: Re-thinking the metarepresentation theory. In S. Baron- Cohen, H. Tager-Flusberg & D. Cohen (Eds.). Understanding Other Minds: Perspectives from Autism (pp. 83- 111). Oxford: Oxford Medical Publications.

Perner, J. & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? Infant and Child Development, 11, 93-105.

Perner, J., Stummer, S. & Lang, B. (1999). Executive functions and theory of mind: cognitive complexity or functional dependence? In P.D. Zelazo, J.W. Astington and D.R. Olson (Eds.). Developing Theories of Intention: Social Understanding and Self Control. Mahwah, N.J.: Erlbaum.

Russell, J. (1996). Agency: Its Role in Mental Development. Hove, U.K.: The Psychology Press.

Russell, J. (Ed.) (1998a). Autism as an Executive Disorder. Oxford: Oxford University Press.

Russell, J. (1998b). How an executive disorder can bring about an inadequate 'theory of mind'. In J. Russell (Ed.) Autism as an Executive Disorder. Oxford: Oxford University Press.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13, 103-128.

Zaitchik, D. (1990). When representations conflict with reality: the pre-schooler's problem with false beliefs and 'false' photographs. Cognition, 35, 41-68.

Zelazo, P.D., Burack, J.A., Benedetto, E. & Frye, D. (1996). Theory of mind and rule use in individuals with Down syndrome: A test of the uniqueness and specificity claims. Journal of Child Psychology and Psychiatry, 37, 479-484.

Zelazo, P.D., Burack, J., Boseovski, J., Jacques, S., & Frye, D. (2001). A cognitive complexity and control framework for the study of autism. In J.A. Burack, T.Charman, N. Yirmiya & P.R. Zelazo Eds.) The Development of Autism: Perspectives from Theory and Research. Mahwah, N.J.: Erlbaum.

Zelazo, P.D. & Frye, D. (1998). Cognitive complexity and control: II. The development of executive control in childhood. Current Directions in Psychological Science, 47, 121-126.

Zelazo, P.D., Jacques, S., Burack, J.A. & Frye, D. (2002). The relation between Theory of Mind and rule use: Evidence from persons with autism-spectrum disorders. Infant and Child Development, 11, 171-195.

Ziatas, K,. Durkin, K, & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. Journal Of Child Psychology and Psychiatry,39, 755-763.

Appendix

Sally-Anne Test Used In Experiments 1 and 2

Here we have two dolls. One is called Sally and the other, Anne. This is Sally and this is Anne.

Can you show me Sally and show me Anne?

Sally puts her marble in the box, and then she goes out for a walk. While she is out Anne takes the marble from the box and moves it over to the basket.

Now Sally has come back from her walk.

Sally was then brought back into the scene and made to stand at an equal distance from the box and the basket.

Test Question: Where will Sally look for her marble?

Reality Question: Where is the marble now?

Memory Question: Where did Sally leave her marble at the start?

Train Task

Training phase (Experiments 1 and 2).

Here is a model airport. Let me show you how it works. Here, we have two landing pads, a blue one and a yellow one. These are where the plane lands – like this.

The experimenter makes a plane land on each pad in turn.

Can you show me the blue pad. Now show me the yellow one. Down here we have a train that takes things from the plane to the train shed. This train is automatic. It doesn't have a driver. There is no-one to see where the plane has landed, so there is a signal here…

Experimenter points to signal lights

…to show the train where to go. When a plane lands on the blue pad, the blue light comes on. That means that the train goes to the blue pad, like this.

Experimenter makes the train go to the blue pad and repeats the demonstration for the yellow pad/light.

Now the plane lands on the yellow pad. See - the yellow light comes on.

Where will the train go to?

Test phase (Experiment 1).

If the child responds correctly on at least five of the training trials, the experimenter presses a switch to change the contingency between the landing pads and the lights.

Now the plane comes in to land on the yellow pad. Oh! Look! The blue light has come on. The plane's on the yellow pad but the blue light is on.

The experimenter then switches off the lights and covers up the landing pads.

Test Question: Where will the train go to?

Reality Question: Where is the plane now?

Memory Question: What colour did the light show?

Test phase (Experiment 2).

As in Experiment 1, testing continues only if the child responds correctly on at least five of the training trials.

Now the plane comes in to land on the yellow pad. Oh! Look! Here comes a bird – it's landing on the yellow pad! See, the yellow light has come on. The plane will have to go to the blue pad. Now the plane's on the blue pad but the yellow light is on.

The experimenter then switches off the lights and covers up the landing pads.

Test, reality and memory questions as in version for Experiment 1.

Author Notes

Earlier versions of this work were presented at the Conference of the Society for Research in Child Development, Indianapolis, March, 1995 and the conference of the ISSBD, Québec, August, 1996. The authors would like to thank Charles Legg, Jake Burack and Sue Leekam for comments, and Jackie Kingston, Shari Joseph and Janet Boseovski for help with data collection. Thanks are also due to the schools and children who took part. The work was supported by grant number 38018/Z/93 from the Wellcome Trust to the first author.

Table 1

Comparative analysis of false belief and mechanical analogue tasks.

	Sally-Anne Task	Train Task
Agent	Sally	Train
Goal	Marble	Plane
Signal	Sally's belief about location of marble	Light indicating location of plane
"False belief" situation	Sally holds wrong belief about location of marble	Signal light wrongly indicates location of plane

Table 2

Chronological and verbal mental ages of the three groups of children who participated in Experiment 1.

	Chronological Age (Months)			BPVS[1] Age (Months)		
Group	M	SD	Range	M	SD	Range
Typically Developing (N = 22)	50.86	4.89	41- 57	50.41	9.97	30- 67
With Intellectual Disabilities (N = 21)	154.38	26.46	103-213	57.10	13.64	40- 92
With Autism	127.13	40.52	71-213	61.21	27.43	35-156

Table 2

Chronological and verbal mental ages of the three groups of children who participated in Experiment 1.

	Chronological Age (Months)			BPVS[1] Age (Months)		
Group	M	SD	Range	M	SD	Range
Typically Developing (N = 22)	50.86	4.89	41- 57	50.41	9.97	30- 67
With Intellectual Disabilities (N = 21)	154.38	26.46	103-213	57.10	13.64	40- 92
With Autism	127.13	40.52	71-213	61.21	27.43	35-156

Table 3

Performance of children from the three groups on both tasks in Experim

Test

	Sally-Anne		Train	
	Fail	Pass	Chooses Plane	Choose Light
Group				
Typically Developing	11	11	12	10
With Intellectual Disabilities	11	10	12	9
With Autism	16	8	15	9

Typically Developing vs With Autism: Sally Anne χ^2 (1, N = 46) = 0.72, ns; Tra 46) = 0.06, ns

Typically Developing vs With Intellectual disabilities: Sally-Anne χ^2 (1, N = 43) = Train χ^2 (1, N = 43) = 0.00, ns

With Autism vs With Intellectual disabilities: Sally-Anne χ^2 (1, N = 45) = 0.45, n 45) = 0.00, ns

Table 4:

Relations between performance on the two tasks in Experiment 1 for the three groups of children.

		Train	
		Chooses Plane	Chooses Light
Group	Sally-Anne		
Typically Developing[a]	Fail	10	1
	Pass	2	9
With Intellectual Disabilities[b]	Fail	10	1
	Pass	2	8
With Autism[c]	Fail	13	3
	Pass	2	6

[a] Significant correlation between measures ($\Phi = 0.73$, $p < 0.001$), no significant difference in difficulty between measures (McNemar: ns).

[b] Significant correlation between measures ($\Phi = 0.72$, $p < 0.002$), no significant difference in difficulty between measures (McNemar: ns)

[c] Significant correlation between measures ($\Phi = 0.55$, $p < 0.006$), no significant difference in difficulty between measures (McNemar: ns).

Table 6

Performance of children from the two groups on both tasks in Experimer

	Test			
	Sally-Anne[a]		Train[b]	
Group	Fail	Pass	Chooses Plane	Chooses Light
Typically Developing	9	16	10	15
With Autism	9	9	7	11

[a] No significant difference between groups (χ^2 (1, N = 43) = 0.37, ns)

[b] No significant difference between groups (χ^2 (1, N = 43) = 0.00, ns)

Table 7

Relations between performance on the two tasks in Experiment 2 for the two groups of children.

Group	Sally-Anne	Train: Chooses Plane	Train: Chooses Light
Typically Developing[a]	Fail	7	2
	Pass	3	13
With Autism[b]	Fail	7	2
	Pass	0	9

[a] Significant correlation between measures ($\Phi = 0.58$, $p < 0.005$), no significant difference in difficulty between measures (McNemar: ns).

[b] Significant correlation between measures ($\Phi = 0.80$, $p < 0.001$), no significant difference in difficulty between measures (McNemar: ns)

Figure 1

Apparatus used in Experiment 1.