White, S.J., Hill, Elisabeth L., Happe, F.G.E. and Frith, U.

Revisiting the strange stories: revealing mentalizing impairments in autism.

Originally published in Child Development

CHILD DEVELOPMENT

Child Development, July/August 2009, Volume 80, Number 4, Pages 1097–1117

Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism

Sarah White
University College London

Elisabeth Hill
University of London

Francesca Happé
King's College London

Uta Frith
University College London

A test of advanced theory of mind (ToM), first introduced by F. Happé (1994), was adapted for children (mental, human, animal, and nature stories plus unlinked sentences). These materials were closely matched for difficulty and were presented to forty-five 7- to 12-year-olds with autism and 27 control children. Children with autism who showed ToM impairment on independent tests performed significantly more poorly than controls solely on the mental, human, and animal stories with greatest impairment on the former and least on the latter. Thus, a mentalizing deficit may affect understanding of biologic agents even when this does not explicitly require understanding others' mental states.

Since Baron-Cohen, Leslie, and Frith's (1985) study, there has been an immense interest in the theory-of-mind (ToM) theory of autism. Individuals with autism were shown to have specific problems in understanding other people's mental states, an ability that has come to be known as mentalizing. Deficits in mentalizing make intuitive sense of the characteristic social and communicative impairments in autism spectrum disorders (ASDs), and relations have been demonstrated between tests of mentalizing and everyday social insight (Frith, Happé, & Siddons, 1994). However, tests to assess mentalizing ability are scarce and the simple tests often used tend to suffer from ceiling effects.

The Strange Stories (Happé, 1994) provide a means of testing advanced mentalizing ability, which is suitable for both higher functioning children and adults (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Gillott, Furniss, & Walter, 2001; Jolliffe & Baron-Cohen, 1999a; Kaland et al., 2005), as well as patients with acquired brain lesions (Happé, Brownell, & Winner, 1999). Subjects read short vignettes and are asked to explain why a character says something that is not literally true. Successful performance thus requires attribution of mental states such as desires, beliefs or intentions, and sometimes higher order mental states such as one character's belief about what another character knows.

Happé's (1994) original set of 24 stories was accompanied by a smaller set of six control stories requiring understanding of physical states. All participants in Happé's study performed at ceiling on these control stories, which were not equated for difficulty with the mental state stories. It is therefore possible that the poor performance of the autism group on the mental state questions could have been due to general comprehension problems or the need to integrate information and make inferences across the text, which were not revealed by the easier physical state stories. In fact, comprehension problems and problems with inferences are a common feature of high-functioning individuals on the autism spectrum (Jolliffe & Baron-Cohen, 1999b; Norbury & Bishop, 2002). Problems in integrating information, one aspect of so-called ''weak central coherence,'' are therefore an alternative explanation for failure on the mental state stories, a possibility discussed by both Happé and Jolliffe and Baron-Cohen (1999a).

In subsequent development of the Strange Stories for functional brain imaging using PET (Fletcher et al., 1995), a subset of the eight most

We would like to thank Mariam Aljunied, Philip Angell, Karolina Jankowska, Hannah Roche, Suman Saha, and Gabriella Worth for help with data collection. Many thanks also go to all the children and adults who participated in this study, and the many schools who so willingly involved themselves. This research was funded by Medical Research Council Grants G78/8085 (S.W.) and G9617036 (U.F.) and joint Medical Research Council/Economic and Social Research Council Grant PTA-037-27-0107 (S.W.).

Correspondence concerning this article should be addressed to Sarah White, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom. Electronic mail may be sent to s.white@ucl.ac.uk.

demanding mental state stories was selected, involving understanding of double bluff, white lie, persuasion, and misunderstanding. To test for the specificity of any difficulty with these stories, eight control stories were created that required reasoning about physical states only, and were of comparable difficulty in healthy young adults. An additional set of eight passages of unlinked sentences was also created; here, participants were required to recall a specific fact from one sentence. All sets therefore involved people and required attention to sentence meaning, memory, and question answering, whereas the mental and physical state sets also required the integration of information between sentences and inference from implicit information, and only the mental state set required mentalizing.

These well-controlled stimuli enabled a ''mentalizing network'' in the brain to be revealed, including medial prefrontal cortex, superior temporal sulcus (STS) and the temporal poles. Reduced activation in this network was seen in adults with Asperger syndrome (AS; Happé et al., 1996). Here, 5 adults were chosen to be extremely high functioning and to perform well at these tasks, to allow valid scan comparisons. Despite this, the behavioral results still showed a significant difference between the groups on the mental state stories but not on the physical or unlinked stories. However, this study had low power to detect even small or moderate effects, and examination of the group means revealed that the performance of the adults with AS was slightly lower than controls on the physical state stories and slightly higher on the unlinked sentences. A recent study of 20 high-functioning children with ASDs (Brent, Rios, Happé, & Charman, 2004), found a similar pattern of performance in a reduced subset of these mental and physical state story sets, showing significantly poorer performance than age- and IQ-matched controls on the mental state set, and a slight trend toward poorer performance on the physical state set.

The present study was motivated by the need to investigate the scope and limitation of the three sets of Strange Stories that were used in the identification of the brain's mentalizing system. The first pilot study therefore tested the three sets in the largest samples of both adult and child high-functioning populations to date. Of particular interest was whether populations with autism, despite a high intellectual level, would have difficulties not only with the mental state stories but possibly also with the physical state stories, either due to more general comprehension problems or to the fact that both types of stories contained human protagonists and were thus not as distinct as they might be in terms of the presence and absence of the requirement to mentalize (see also Saxe & Kanwisher, 2003).

Pilot Study With Story Materials Used by Fletcher et al. (1995)

Method

Participants

Twenty-three high-functioning adults and 39 children, all clinically diagnosed as having an autistic disorder (including autism, AS, and ASD), took part in this study, compared to 40 adult and 41 child control participants. The groups did not significantly differ on gender: adults, $\chi^2(1) = 2.80$, $p = .094$; children, $\chi^2(1) = 2.94$, $p = .086$; age: adults, $t(61) = 1.29$; children, $t(78) = 0.33$; or nonverbal (performance IQ): adults, $t(61) = 0.41$; children, $t(64.62) = 0.03$. The majority of participants were Caucasian and from middle-class families. Although there was a trend for more males to be present in the ASD than control groups, no differences were found between males and females on any matching or experimental measure in either control group, indicating that gender could not account for any group differences.

The adults were all selected to have nonverbal IQs above 85 and were aged between 16 and 63 years. The children were aged between 7 and 11 years and had nonverbal IQs ranging from 70 upward. As can be seen in Table 1, the children with ASD had a lower mean verbal IQ than the nonverbal IQ-matched controls, $t(78) = 5.42$, $p < .001$, whereas the ASD adults were well matched to the control adults on verbal IQ, $t(61) = 0.50$.

The children were recruited through mainstream schools and special schools for children with such diagnoses. Adult participants were recruited through advertisements placed around the local community and through AS support groups throughout the United Kingdom. The ASD adults had significantly elevated scores on the Autism-Spectrum Quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), $t(48) = 9.09$, $p < .001$; possible range of scores 0–50. None of the control participants were reported or were known to have any developmental disorders or family history of such difficulties.

Materials

The materials were the three text sets used first in Fletcher et al.'s (1995) study: eight mental state

Table 1
Participant Characteristics and Background Theory of Mind Results From the Pilot Study

	ASD children	Control children	ASD adults	Control adults
N (male:female)	39 (33:6)	41 (28:13)	23 (17:6)	40 (21:19)
Age (years)	9.73 (1.22)	9.64 (1.26)	31.22 (13.32)	35.50 (12.30)
Nonverbal IQ	98 (18.34)	98 (11.88)	114 (18.25)	111 (16.11)
Verbal IQ child***	86 (10.84)	99 (9.91)	107 (12.50)	109 (13.79)
Clinical diagnosis	7 Autism 11 AS 21 ASD	—	23 AS	—
AQ (cutoff 32) adult***	—	—	34.14 (7.62)	16.62 (7.85)
Sally-Ann (% passing) child*	72	93	96	100
Smarties (% passing) child**	74	97	—	—
Ice Cream Van (% passing) child***	41	90	—	—
Birthday Puppy (% passing) child***	58	97	—	—
Coat Story (% passing) adult**	—	—	70	95

Note. Values are given as mean (standard deviation) except when otherwise stated. The asterisks following ''adult'' or ''child'' indicate which pairing of groups are significantly different from one another. ASD = autism spectrum disorder; AS = Asperger syndrome; AQ = autism-spectrum quotient.
*$p < .05$. **$p < .01$. ***$p < .001$.

stories, eight physical state stories, and eight passages of unlinked sentences (see the Appendix for the full test materials).

Procedure

Informed consent (from the participant or their parent, as appropriate) was obtained prior to involvement in the study. Each participant was tested individually in a quiet room either at the UCL Institute of Cognitive Neuroscience (for the vast majority of adults), at their school (for the vast majority of children), or at their home.

For the children, verbal IQ was measured using the British Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton, & Burley, 1997) whereas Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1988) were used to ascertain nonverbal IQ. The intelligence of the adult participants was measured using the Wechsler Adult Intelligence Scales (WAIS–III–UK; Wechsler, 1997).

The children performed four additional tests of ToM: two first-order false belief tasks, Sally-Ann (Baron-Cohen et al., 1985) and Smarties (Perner, Leekam, & Wimmer, 1987), and two second-order false belief tasks, Ice Cream Van (Baron-Cohen, 1989) and Birthday Puppy (Sullivan, Zaitchik, & Tager-Flusberg, 1994). The adults performed both the Sally-Ann task and a second-order Coat Story task (Bowler, 1992).

All three sets of text (Fletcher et al., 1995) were presented on cards to each participant in three separate blocks, randomized both within each story type as well as between story type. These were read aloud to the children, whereas the adults read the stories silently themselves. When the story had been read, the card was then turned over and the participant was presented with a question about its content. Verbal responses were recorded and the accuracy of each response was rated on a 0–2 scale by two researchers: 0 for an incorrect answer, 1 for a partially or implicitly correct answer, and 2 for a fully correct answer. A maximum score of 16 was therefore possible. Scoring schemes for the stories are given in the Appendix. Good agreement was reached (intraclass correlation coefficient = .89) between the first author and a corater blind to group. Response time was also recorded for the adult participants, defined as the time from starting to read the story to the time when starting to answer verbally the question. Mean response times across all story responses in a set were calculated.

Results

Children

Significantly, more children in the ASD group than the control group failed each of the false belief tests: Sally-Ann, $\chi^2(1) = 6.04$, $p = .014$; Smarties, $\chi^2(1) = 8.57$, $p = .003$; Ice Cream Van, $\chi^2(1) = 21.66$, $p < .001$; Birthday Puppy, $\chi^2(1) = 17.35$, $p < .001$.

Figure 1 shows the results from the stories tasks. As the groups were not matched for verbal ability

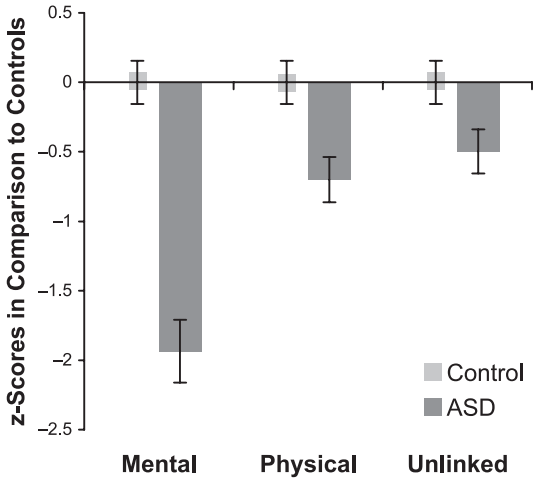

Figure 1. Performance on mental and human physical state stories and unlinked sentences by control and autism spectrum disorder (ASD) children after accounting for verbal level in the pilot study. Scores are presented as *z* scores relative to the control group; error bars are standard errors.

and story performance was generally related to verbal ability ($r \geq .40$, $p \leq .009$ except for the control group on the unlinked sentences: $r = .14$), individual performance levels across each story set were calculated by entering data for each story type from the control group as the dependent variable in a regression with verbal ability (BPVS raw score) as the predictor variable. This regression equation was then applied to the ASD group and residuals were collected, and were converted to *z* scores in relation to the control group's mean and standard deviation. These *z* scores were then used in all further analyses. Given the control group range of verbal ability spanned that of the ASD group, this method was considered appropriate (Miller & Chapman, 2001).

An analysis of variance (ANOVA) comparing story type (mental/physical) by group (ASD/control) revealed a main effect of group, $F(1, 78) = 37.89$, $p < .001$, $\eta_p^2 = .32$, with the ASD group performing below the control group. An interaction was also found between group and story type, $F(1, 78) = 22.95$, $p < .001$, $\eta_p^2 = .23$; the main effect of story type was identical to this interaction as *z* scores relative to the control group mean were used. Post hoc comparisons revealed that the autism group performed below the controls on both the mental ($p < .001$, $\eta_p^2 = .39$) and physical ($p = .003$, $\eta_p^2 = .23$) stories, and that the ASD group also performed more poorly on the mental than physical stories ($p < .001$, $\eta_p^2 = .50$).

As the unlinked sentences had a ''yes–no'' rather than an open question, these scores were analyzed separately. A *t* test revealed that the performance of the children with autism was lower than the controls on these passages, $t(78) = 2.23$, $p = .029$, $d = .50$.

Adults. There was no group difference in the number of adults passing the first-order Sally-Ann task, $\chi^2(1) = 1.77$, but significantly more ASD adults failed the second-order Coat Story task, $\chi^2(1) = 7.72$, $p = .005$.

An ANOVA comparing story type (mental, physical) by group (ASD, control) revealed a borderline main effect of group, $F(1, 61) = 3.97$, $p = .051$, $\eta_p^2 = .06$, due to the ASD adults performing less accurately than the control group (see Figure 2a). A main effect of story type was also found, $F(1, 61) = 7.64$, $p = .008$, $\eta_p^2 = .11$, indicating that the mental state stories were easier than the physical state stories. However, there was no interaction between group and story type, $F(1, 61) = 0.31$. Additionally, no difference in performance was found between groups on the unlinked sentences, $t(60) = 0.01$.

However, when the same analysis was repeated using the response time data, in addition to main effects of group, $F(1, 60) = 42.45$, $p < .001$, $\eta_p^2 = .41$, and story type, $F(1, 60) = 5.67$, $p = .020$, $\eta_p^2 = .09$, a trend toward an interaction between group and story type was found, $F(1, 60) = 3.56$, $p = .064$, $\eta_p^2 = .06$. Both groups were quicker at responding to the mental than physical state stories (ASD $p = .009$, $\eta_p^2 = .29$; controls $p < .001$, $\eta_p^2 = .56$). However, although there was a significant difference between the groups on the mental state stories ($p = .017$, $\eta_p^2 = .14$) with the ASD group responding slower, this difference was not significant on the physical state stories (see Figure 2b). There was also a difference between the groups on the unlinked sentences ($p = .015$, $\eta_p^2 = .14$).

Discussion

The results from the standard false belief tasks confirm previous studies demonstrating the presence of specific mentalizing deficits at the group level in both children and adults. This is remarkable as the performance of the present high-functioning participants was close to ceiling. As regards the Strange Stories, close to ceiling performance probably precluded differences in accuracy to be shown in the adult group on the mental state set (74% of ASD group and 85% of control group scoring 12 or more of 16), but the ASD group was significantly slower in responding only to mental state stories specifically. In the ASD child sample,

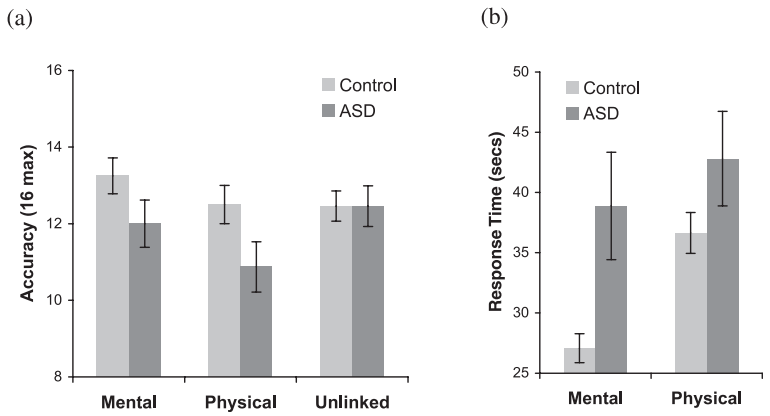

Figure 2. Performance on mental and human physical state stories and unlinked sentences by control and autism spectrum disorder (ASD) adults in the pilot study. Accuracy is shown in (a) and response times in (b); error bars are standard errors.

performance was generally at a lower level compared to controls, but the greatest impairment was revealed on the mental state stories. Nevertheless, the small if significant differences on the nonmental state stories suggest that other confounding factors may be inherent in the story materials, and these might be identified by systematically changing story content.

One interesting possibility is that differences between mental and physical state story performance might have been diminished as both require processing information about human characters. For example, a study by Blair, Frith, Smith, Abell, and Cipolotti (2002) showed that visual recognition memory in high-functioning adults with autism was significantly poorer for stimuli that were potential agents capable of self-initiated movement, compared to buildings or leaves that could not be described as self-propelled agents. Likewise, the recognition of biologic motion from random dot kinematograms has been shown to be impaired in children with ASD (Blake, Turner, Smoski, Pozdol, & Stone, 2003). In line with these behavioral findings, the STS, the main brain area thought to process biologic motion (Bonda, Petrides, Ostry, & Evans, 1996) has been shown to be functionally and structurally abnormal in individuals with ASD (Boddaert et al., 2004; Pelphrey, Adolphs, & Morris, 2004; Waiter et al., 2004). The STS region is one of the main components of the mentalizing system in the brain and has consistently been found to be activated in typical adults when reading both the mental and physical state Strange Stories (Fletcher et al., 1995; Gallagher et al., 2000). Furthermore, the same neural areas of medial prefrontal cortex, another main component of the brain's mentalizing network, were activated for both animals and humans when volunteers were asked to think about their mental as opposed to physical states (Mitchell, Banaji, & Macrae, 2005). This might indicate that processing information about all animate agents may be impaired in ASD and hence depress performance even in those stories that did not explicitly require mentalizing.

There is also evidence to show that the network of brain areas thought to be involved in mentalizing is also activated when making nonmentalistic semantic judgments about people as opposed to objects (Mitchell, Heatherton, & Macrae, 2002), although possibly to a lesser extent than when performing an explicit mentalizing task (Fletcher et al., 1995). This suggests that mentalizing may aid understanding of agents, even when mental state processing is not explicitly required.

In line with these considerations, two new sets of physical stories for use with children involving either animals or natural events were therefore added to the original three sets of stories. They therefore allow us to investigate the level at which children with ASD have problems, specifically with mentalizing, with making inferences about agents, or with story comprehension (see Table 2). In the next study, mentalizing ability was also independently assessed by a large battery of ToM tests. It was anticipated that, in a group of high-functioning children with ASD, individual differences would be found in mentalizing ability. It was of particular interest whether a severely ToM impaired subgroup would also show poor performance on mental state stories, but unimpaired performance on the other stories.

Table 2
Task Requirements for the Different Strange Story Sets

	Mental	Human	Animal	Nature	Unlinked
Explicit mentalizing	✔	✗	✗	✗	✗
Thinking about humans	✔	✔	✗	✗	✗
Thinking about animate agents	✔	✔	✔	✗	✗
Integration of info across text	✔	✔	✔	✔	✗
Sentence comprehension	✔	✔	✔	✔	✔

As regards testing of children, it should be noted that the mental state stories were substantially easier than the human physical state stories for the control children in the pilot study and that performance on the physical state stories was generally very poor. This is likely to be due to the contents of some of the physical stories requiring a level of world knowledge beyond that of most children of this age to comprehend the story. It is therefore possible that a floor effect may have been acting differentially across story types in the children, reducing the difference between the performance of the two groups on the physical state stories only (although this could not have been the case in the adults' response times). This would mean that the interaction, and therefore the specific mentalizing impairment, may have been an artifact of this floor effect. The original set of human physical state stories was therefore modified to be more appropriate for children and matched to the mental state stories for difficulty, to eliminate any differential floor effects.

The poor performance by the children with ASD on the unlinked sentences was also difficult to interpret as these passages involved closed questions requiring a ''yes'' or ''no'' response and therefore encouraged guessing. To ensure correct answers represented understanding and memory of sentence meaning rather than chance performance, this set was modified to involve open questions (e.g., When is Mary's birthday?).

Main Experiment With Newly Devised Story Materials

This next study attempts to identify and provide a detailed analysis of the difficulties individuals with autism have with understanding the Strange Stories and the possible causes of these. The physical stories and unlinked sentences were modified to make them more appropriate for a child population, and two novel sets of stories were added to aid the investigation of difficulties in understanding mental states and agents and with text comprehension. The procedure was also standardized by prerecording the stories and presenting them on a computer.

Method

Participants

A new sample of 45 children with ASD aged 7–12 years took part, matched to 27 control participants for gender, $\chi^2(1) = 2.51$; age, $t(70) = 0.74$; verbal, $t(70) = 0.95$; and performance IQ, $t(70) = 1.78$. All but 3 children in each group were Caucasian and the majority were from middle-class families. As can be seen in Table 3, the children with ASD had slightly lower mean verbal and performance IQ scores than the controls, so individual variation was taken into account in the analyses. As their IQ scores indicate, these children were an extremely high functioning sample and all but two attended a mainstream school. All these children had previously received a diagnosis of autism, AS, or ASD from a qualified clinician prior to participation in the

Table 3
Participant Characteristics and Background Theory of Mind Results From the Main Experiment

	ASD children	Control children
N (male:female)	45 (41:4)	27 (21:6)
Age (years)	9.24 (1.39)	9.48 (1.35)
Verbal IQ	111 (14.70)	115 (15.78)
Performance IQ	98 (11.18)	103 (12.35)
Clinical diagnosis	8 Autism	—
	25 AS	
	12 ASD	
3Di		
Social (cutoff 10)***	12.71 (4.20)	3.34 (2.00)
Communication (cutoff 8)***	14.77 (3.37)	3.73 (1.68)
Repetitive behavior (cutoff 3)***	4.90 (2.40)	0.35 (0.54)
Background ToM battery (max. 25)**	16.34 (3.80)	18.98 (2.54)

Note. Values are given as mean (standard deviation) except when otherwise stated. ASD = autism spectrum disorder; AS = Asperger syndrome; 3Di = Developmental, Dimensional, and Diagnostic Interview; ToM = theory of mind.
**$p < .01$. ***$p < .001$.

study. In addition, all met objective criteria for an ASD on the Developmental, Dimensional, and Diagnostic Interview (3Di; Skuse et al., 2004) at the time of this study, whereas none of the controls did, social, communication, and repetitive behavior domains, $t(66) > 11$, $p < .001$; possible range of scores: social 0–30, communication 0–26, repetitive behavior 0–12. The 3Di measure is similar to the Autism Diagnostic Interview (ADI–R; Lord, Rutter, & Le Couteur, 1994) with which it correlates highly (Skuse et al., 2004). None of the control participants was reported or was known to have any developmental disorders or family history of such difficulties.

Materials and Procedure

Verbal and performance IQ were measured using the WISC–III–UK; (Wechsler, 1992).

Standard ToM battery. The children performed 12 tests of ToM previously established as discriminating between autistic and typically developing samples of children. They were presented in a random order with accompanying props. Seven of the tasks were those used by Wellman and Liu (2004) and the 5 additional tasks were expected to be more challenging for the older children involved here (Ice cream Van: Baron-Cohen, 1989; Penny Hiding: Baron-Cohen, 1992; Sally-Ann: Baron-Cohen et al., 1985; Interpretational False Belief: Luckett, Powell, Messer, Thornton, & Schulz, 2002; Birthday Puppy: Sullivan et al., 1994). These were scored out of a total of 25.

Eleven of the tasks consisted of situations in which a character and the child had diverse desires or diverse beliefs; where a character had a lack of knowledge, a false belief, or a false belief about another character's false belief; or where a character wanted to create a false belief in others. In all cases, the child was asked to predict the knowledge of, behavior of or emotion felt by another character on the basis of their mental state; this answer was marked as a pass or fail (1 or 0). In all but the three simplest tasks, the child was also asked to justify why they had made that prediction; this was marked as a correct mental state justification (e.g., ''because she doesn't *know* it's in the box''), a correct nonmental state justification (e.g., ''because she left it in the basket''), or an incorrect justification (e.g., ''because no-one stole it apart from Anne''; 1, 0.5, or 0, respectively). This additional scoring aimed to check for false positive responses when the child was achieving the correct answer by guessing, and also was expected to increase the variation in responses with the aim of avoiding ceiling effects. Control questions were administered (e.g., memory and reality questions; also prompt questions during the more complex tasks), which all individuals were required to pass, to check for comprehension of the scenario.

The remaining task, accounting for a maximum of six marks, was the penny-hiding task; this involved six trials when the experimenter modeled hiding the coin to the child, followed by six test trials when the child was encouraged to hide the coin. One point was awarded on each trial if the experimenter was unable to tell which hand the coin was hidden in; half a point was given if the child made no obvious mistake but held their hands asymmetrically; and the child gained no points if they failed to deceive the experimenter (e.g., not hiding both hands behind their back, bringing only one hand out, not closing the empty hand, leaving the coin in sight).

Modified sets of Strange Stories. The three text sets used in the pilot study were again used but with four of the original human-physical stories (last four in Appendix; these were the most difficult according to results from an independent sample of typically developing children) replaced by four new, easier human-physical stories (from here on referred to as human stories; the full new human set was matched in difficulty to the mental state set in an independent sample of typically developing children). The questions accompanying the eight passages of unlinked sentences were modified to be open rather than closed questions. In addition, two new sets of eight animal-physical stories and eight nature-physical stories were created, matched to the mental state set for difficulty in an independent sample of typically developing children (see the Appendix for the full test materials). To illustrate the difference between the story types, an example of each is given as follows.

Mental

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

Q: Why did the prisoner say that?

Human

Bob and Jim are best friends. They are both 10 years old. Bob has brown hair, green eyes and is over 5 feet tall. Jim looks very different to Bob. He has blonde hair and blue eyes and he is much smaller than Bob. Bob and Jim go on an outing to the fun fair. They go on lots of rides. For the last ride of the day they decide to go on the big rollercoaster. But there is a sign which says: For safety reasons, no persons under 5 feet are allowed on.

Q: Why does only Bob go on the rollercoaster?

Animal

Emperor penguins live in the Antarctica, where it is extremely cold. There is always snow on the ground and ice on the surface of the sea. Emperor penguins can often be found standing clumped together in huge, huddled masses. Every few minutes, a penguin in the middle of the huddle moves to the edge of the huddle, changing places with one of the penguins on the outside of the group.

Q: Why do the penguins keep changing places?

Nature

It is late in April and the sky has been clear and blue all morning. An empty fountain sparkles in the middle of the park. As the day goes on, the sun occasionally disappears behind little white fluffy clouds, soon to appear again on the other side. However in the late afternoon, the sky becomes dark and filled with lots of gray clouds. A little after this, the fountain starts to spout water.

Q: Why did the fountain spout water?

Unlinked

At the edge of the road, a little grass was growing. He reaches out to find the light switch. A sailor who has just left his ship is walking to the town. She has to decide where to keep the pasta. At last daylight came, and Tommy got out of bed to open his presents. Jim knows all about investing money, as he works in a large bank. They exchanged a few brief words about the weather.

Q: Why did Tommy get out of bed?

All five sets of Strange Stories text were presented on a laptop computer using E-Prime software, both as written and spoken text to aid concentration. Each story was prerecorded by an adult male speaker and lasted approximately 30 s. The full text of each story was presented on the screen and the child was invited to follow the words while listening to the recording. Five hundred milliseconds after the spoken text had ended, a question about its content was presented on the screen and simultaneously spoken by the prerecorded voice. The experimenter recorded the child's answer verbatim on paper and this was scored later.

The stories were presented in blocks of eight stories, one block for each story type. The animal and nature stories were presented in a second test session approximately 4 months after the other three sets of stories. The order of the blocks within each session and the order of the stories within each block were randomized. The accuracy of each verbal response was rated on a 0–2 scale as before (scoring schemes are given in the Appendix). Good agreement (intraclass correlation coefficient = .81) was reached between the first author and a corater blind to group based on a sample of answers to each question (20% of answers selected randomly for each story).

Results

As can be seen from Table 3, performance on the ToM battery differed between the groups, with the ASD group performing more poorly than the controls, $F(1, 67) = 8.556$, $p = .005$.

Individual performance levels across all story sets were calculated by entering data for each story type as the dependent variable in a regression, with age, verbal, and performance IQ as predictors. Residuals were collected and hence represented individual variance in the story data that could not be explained by age or IQ, and these scores were converted to z scores in relation to the control group's mean and standard deviation. These z scores were then used in all further analyses.

A good match for difficulty was obtained between the new and old story sets in the typically developing children (mean of 11 or 12 of 16 for all story sets; mean of 9 for unlinked passages). A 5×2 repeated measures ANOVA comparing story type by group revealed a main effect of group, $F(1, 70) = 8.59$, $p = .005$, $\eta_p^2 = .11$, with the performance of the ASD group generally falling below that of the controls. Furthermore, a significant interaction was found between story type and group, $F(4, 280) = 2.78$, $p = .027$, $\eta_p^2 = .04$; again, the main effect of story type was identical to this interaction

as *z* scores relative to the control group mean were used, with the ASD group performing below the control group only on the mental ($p = .002$, $\eta_p^2 = .13$), human ($p = .009$, $\eta_p^2 = .09$), and animal ($p = .020$, $\eta_p^2 = .08$) stories. This interaction followed a linear relation between the groups across the different story types, $F(1, 70) = 8.69$, $p = .004$, $\eta_p^2 = .11$, with the greatest difference on the mental state stories and the smallest difference on the unlinked sentences.

Item analyses were conducted to investigate which stories were influencing these group differences. Importantly, across all eight stories in each of the mental, human, and animal sets, the ASD group mean was consistently lower than that of the control group. Significant group differences were found on five of the mental (in the order in the Appendix, stories 1, 2, 4, 5, 6), four of the human (original physical stories 2, 3, 4, new human physical story 1), and three of the animal stories (stories 1, 2, 7), although trends toward group differences were seen on the other stories too.

As a stringent test of the stories' power to assess poor mentalizing ability, the ASD group was split into two subgroups on the basis of their performance on the background ToM battery. Any child with ASD performing more than 1.65 *SD* (fifth percentile) below the corrected control mean (see White et al., 2006, for further details of methodology used here) was defined as having unquestionably poor ToM (15 children) and the remaining children as having relatively good ToM (30 children). This latter group included children whose performance was still below the control group mean (22 children) as well as those performing above the control group mean (8 children). The two subgroups of poor ToM and relatively good ToM did not differ from each other on age, $t(43) = 0.103$; verbal, $t(43) = 1.00$; or performance IQ, $t(43) = 0.68$, but the children with poor ToM were reported to display more severe social, $t(39) = 2.49$, $p = .017$, $d = .83$, and communication, $t(39) = 2.20$, $p = .034$, $d = .69$, symptoms, but not repetitive behaviors, $t(39) = 0.34$, on the 3Di. Although there was overlap between those children with ASD with poor or relatively good ToM, every child with poor ToM as defined by the ToM battery had performance below the control mean on the mental state Strange Stories. Furthermore, the mental state Strange Stories were highly correlated with the ToM battery in the ASD group ($r = .42$, $p = .001$); as variance in age and IQ had already been accounted for in these variables, this relation is independent of these factors.

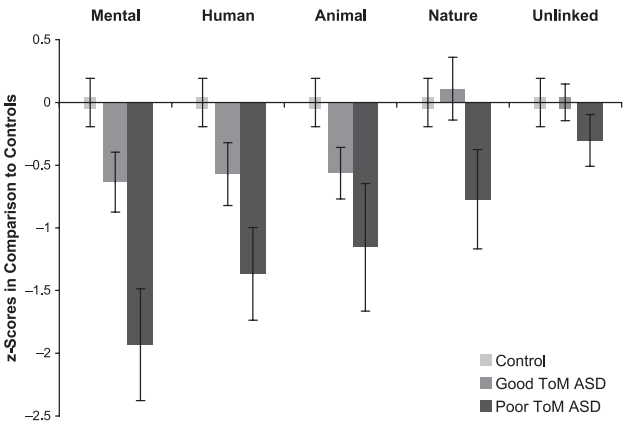

Figure 3. Performance on story sets by control and autism spectrum disorder (ASD) children after accounting for age and IQ in the main experiment. The ASD group is divided into those with relatively good and poor theory of mind (ToM) on the basis of performance on the background ToM tests. Scores are presented as *z* scores relative to the control group; error bars are standard errors.

A 5 × 3 repeated measures ANOVA comparing story type by group (poor ToM ASD, relatively good ToM ASD, control) was performed. As before, this produced main effects of group, $F(2, 69) = 9.77$, $p < .001$, $\eta_p^2 = .22$, as well as an interaction, $F(8, 276) = 2.01$, $p = .045$, $\eta_p^2 = .06$, which followed a linear relation, $F(2, 69) = 5.89$, $p = .004$, $\eta_p^2 = .15$, between the groups across the different story types (see Figure 3). Post hoc tests revealed that the poor ToM ASD group was performing overall more poorly than both other groups ($p \leq .008$, $\eta_p^2 = .15$), which were performing similarly. The linear interaction resulted mainly from the poor ToM subgroup performing below the controls on the mental ($p < .001$, $\eta_p^2 = .35$), human ($p = .001$, $\eta_p^2 = .25$), and animal ($p = .015$, $\eta_p^2 = .14$) stories. The ASD subgroup with relatively good ToM did not differ significantly from the controls in their performance. The relatively good and poor ToM subgroups differed significantly in their performance on the mental stories only ($p = .007$, $\eta_p^2 = .16$).

Discussion

The five sets of stories used in this experiment were found to be well matched in difficulty for the children involved, indicating that any interaction of story type by group could not be accounted for by floor or ceiling effects acting differentially across the different story sets. This allows for greater confidence in the results than for those of the pilot study, although the results from both studies were very similar.

The results from the mental state stories clearly indicate that children with ASD and poor ToM, as indicated by an independent test battery, have significant problems both in comparison to typically developing children and children with ASD who have relatively good ToM. Importantly, their poor performance on the mental state stories is likely to indicate a lack of understanding of mental states rather than a problem comprehending text. The strong correlation between the ToM battery and the mental state stories also provides a validity check that both instruments are tapping into the same underlying ability. This is consistent with findings from neuroimaging studies that these mental state stories activate the brain's mentalizing network. A future battery of ToM tests designed along psychometric principles would thus benefit from the inclusion of the mental state Strange Stories.

The present study is the first time that text comprehension not involving agents has been investigated in autism. In the case of the nature stories, no impairment was shown in the ASD group, indicating intact text comprehension abilities. In addition, the different types of stories allowed examination of story understanding when different degrees of mentalizing may have been evoked. Difficulties with the other sets, however, cannot be attributed to more general text comprehension difficulties.

Analysis of the interaction between story type and group revealed a linear relation in performance differences between the groups across the story types (see Figure 3). First, this indicated that it was the children with poor ToM who were having difficulty with the stories and therefore that poor ToM abilities might result in difficulty understanding not only the mental but also the human and animal sets. In addition, the linear interaction suggested that each consecutive set of stories, which were designed to be one more conceptual step away from the mental state stories (see Table 2), was slightly less difficult for the ASD group with poor ToM than the previous set. In other words, the greater the opportunity for mentalizing during the task, the more impaired the poor ToM group appeared. This is reminiscent of the trend seen in previous studies, and found in the pilot study, for the mental state stories to be more difficult for individuals with ASD than the human physical stories.

It seems likely that the stringent definition of poor ToM in the ASD group resulted in some children with ToM difficulties being categorized in the relatively good ToM group; the interaction effect size in the main experiment was small, most probably due to the group of children with relatively good ToM performing between the controls and the children with poor ToM on three of the measures for this reason. This is supported by the fact that these three measures were the mental, human, and animal stories, exactly those that were difficult for the children with poor ToM.

One parsimonious way of interpreting the interaction results is that the human stories may elicit mentalizing in typically developing individuals despite no explicit need to do so; individuals with autism and poor ToM skills therefore perform poorly on these tasks. This is consistent with previous findings of nonmentalistic text comprehension difficulties in autism (Dennis, Lazenby, & Lockyer, 2001; Jolliffe & Baron-Cohen, 1999b; Norbury & Bishop, 2002; Ozonoff & Miller, 1996), all of which used tasks involving humans' actions, words, or thoughts. A similar implication from the results is that typically developing individuals anthropomorphize nonhumans, as in the animal stories, to understand their actions in terms of mental states; again, individuals with autism and poor ToM skills therefore perform poorly on these tasks. Previous work from two independent groups of researchers supports this notion, as control participants appear to attribute mental states to animated triangles instinctively and with ease, whereas individuals with autism show difficulties in this area (Abell, Happé, & Frith, 2000; Castelli, Frith, Happé, & Frith, 2002; Klin, 2000). An alternate possibility is that understanding agents is one component of mentalizing and therefore that difficulties understanding agents might contribute toward the mentalizing difficulties seen in autism.

Conversely, the nature stories and unlinked sentences do not elicit mentalizing as they do not involve understanding of agents; instead, these story sets revealed that both text and sentence comprehension in the materials presented here are intact in children with ASD, regardless of ToM ability. Although the passages of unlinked sentences do involve humans, the questions did not require the children to understand the actions of these people (see Table 2). This implies that the text comprehension difficulties and mentalizing difficulties both commonly reported in autism have hitherto been confounded due to the materials used. Here, where the two abilities are separated, it seems that only mentalizing is a problem in autism, not text comprehension. Mentalizing difficulties may therefore have been mistakenly regarded as text comprehension difficulties in the past.

One possible limitation is in the presentation of different story types during different sessions. Although the direct comparison of the animal and nature stories with the other three story sets should be treated with slight caution given they were administered approximately 4 months apart, it seems unlikely that this would have a differential effect on the results across the different groups.

General Discussion

The present results provide fresh evidence for a specific deficit in mentalizing ability in ASD, supporting the well-established ToM account. Use of additional new story sets in high-functioning children revealed that this impairment affects comprehension of stories involving people and animals. This may be because mentalizing is also evoked to a lesser degree by stories that do not explicitly require understanding of others' mental states and even less still by stories involving nonhuman agents. Alternatively, reasoning about biologic agents may pose a difficulty for children with ASD that contributes toward their mentalizing impairment. Importantly, we could rule out a general comprehension problem as the same children who had difficulty making inferences about human and animal actions were able to make inferences about natural events, revealing intact text and sentence-level comprehension.

This finding would be further supported by replication with all five story sets in adults and further investigation of the neural mechanisms supporting mental state reasoning across these story sets in autism. Following from the hypothesis that mentalizing is evoked to a lesser degree by the human and animal stories, we would predict that a neuroimaging study using these stories would demonstrate a gradual decrease in activity in the network of brain areas associated with mentalizing across the story types from mental to nature in controls, and a decrease in the difference in this neural activity between controls and individuals with autism across these story sets. This is consistent with studies showing an overlap of regions involved in the processing of biologic agents with those in the mentalizing network and that these areas are activated to a lesser extent by tasks involving agents than when considering mental states (Fletcher et al., 1995; Gallagher et al., 2000; Mitchell et al., 2002, 2005; Saxe & Kanwisher, 2003; Wheatley, Milleville, & Martin, 2007).

One issue that merits further discussion is that of heterogeneity within the ASD groups. Although we are confident that there are robust group differences in mentalizing as found in a large number of studies, we do not claim that every individual with ASD displays the same degree of mentalizing problems with the present tests. In the main experiment here, a subgroup of children with ASD had distinct and severe ToM impairments (with performance below the fifth percentile for the control group), whereas a few children with ASD appeared to be performing complex mental state tasks at a level commensurate with their age and general ability, although the vast majority fell below the control mean (82%). Whether good ToM task performance in this minority of children is due to some form of compensatory learning or indicates intact mentalizing abilities remains to be investigated, possibly through a neuroimaging study of children with autism with good ToM task performance; normal activation of the mentalizing network would be expected if mentalizing abilities were truly intact. Severe ToM impairments also seemed to reflect more severe social and communication difficulties, supporting the role of ToM impairment in the presence of autistic symptomatology. All the children in the ASD group did display significant symptoms of autism though, either upholding the idea of compensation in the children with good ToM performance or indicating that there may be an alternative cause of their symptoms.

This study confirms the value of the Strange Stories in revealing difficulties in mental state reasoning; particularly, these tasks seem better at capturing individual variance in this ability than traditional ToM tests. As the modified stories proved suitable for use with high-functioning children with autism, we would recommend the use of the new sets of stories as they are well matched to the mental state stories in difficulty. The nature stories in particular, may act as an ideal comparison set, controlling for text comprehension abilities. As regards testing of high-functioning adults, it is clear that a new set of stories that allows for response time data to be collected would be a logical next step.

Given the success of the Strange Stories in conjunction with brain imaging studies, the modified stories would seem to provide a conceptually neater method for future studies of ToM abilities in autism. The separation of text comprehension, biologic agency and an explicit requirement for mentalizing would allow a powerful paradigm for the delineation of relevant brain regions. Furthermore, the comparison of those individuals with autism who largely succeed on ToM tasks and those

who largely fail, despite equal intelligence and age, would be revealing with regards individual differences in the underlying causes of autism. In the future, it may therefore be possible to use the stories to identify ''fine cuts'' between superficially similar but essentially different components of social cognition.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*(2), 285–297.

Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry*, *33*(7), 1141–1155.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a ''theory of mind''? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Blair, R. J., Frith, U., Smith, N., Abell, F., & Cipolotti, L. (2002). Fractionation of visual memory: Agency detection and its impairment in autism. *Neuropsychologia*, *40*(1), 108–118.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, *14*(2), 151–157.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., et al. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: A voxel-based morphometry MRI study. *Neuroimage*, *23*(1), 364–369.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, *16*(11), 3737–3744.

Bowler, D. M. (1992). ''Theory of mind'' in Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *33*(5), 877–893.

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism*, *8*(3), 283–299.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*(8), 1839–1849.

Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*, *31*(1), 47–54.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale (BPVS)* (2nd ed.). London: NFER-Nelson.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition*, *57*(2), 109–128.

Frith, U., Happé, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development*, *3*, 108–124.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of ''theory of mind'' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21.

Gillott, A., Furniss, F., & Walter, A. (2001). Anxiety in high-functioning children with autism. *Autism*, *5*(3), 277–286.

Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154.

Happé, F., Brownell, H., & Winner, E. (1999). Acquired ''theory of mind'' impairments following stroke. *Cognition*, *70*(3), 211–240.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). ''Theory of mind'' in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport*, *8*(1), 197–201.

Jolliffe, T., & Baron-Cohen, S. (1999a). The Strange Stories test: A replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, *29*(5), 395–406.

Jolliffe, T., & Baron-Cohen, S. (1999b). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: Is local coherence impaired? *Cognition*, *71*(2), 149–185.

Kaland, N., Moller-Nielsen, A., Smith, L., Mortensen, E. L., Callesen, K., & Gottlieb, D. (2005). The Strange Stories test—A replication study of children and adolescents with Asperger syndrome. *European Child and Adolescent Psychiatry*, *14*(2), 73–82.

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The Social Attribution Task. *Journal of Child Psychology and Psychiatry*, *41*(7), 831–846.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview–Revised: A revised version of a

diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Luckett, T., Powell, S. D., Messer, D. J., Thornton, M. E., & Schulz, J. (2002). Do children with autism who pass false belief tasks understand the mind as active interpreter? *Journal of Autism and Developmental Disorders*, *32*(2), 127–140.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, *28*(4), 757–762.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15238–15243.

Norbury, C. F., & Bishop, D. V. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language and Communication Disorders*, *37*(3), 227–251.

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language*, *52*(3), 411–434.

Pelphrey, K., Adolphs, R., & Morris, J. P. (2004). Neuroanatomical substrates of social cognition dysfunction in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(4), 259–271.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three year olds' difficulty with false belief: The case for conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–127.

Raven, J., Court, J., & Raven, J. (1988). *Raven's Standard Progressive Matrices*. London: J.C. Raven.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in ''theory of mind.'' *Neuroimage*, *19*(4), 1835–1842.

Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., et al. (2004). The developmental, dimensional and diagnostic interview (3di): A novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(5), 548–558.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, *30*(3), 395–402.

Waiter, G. D., Williams, J. H., Murray, A. D., Gilchrist, A., Perrett, D. I., & Whiten, A. (2004). A voxel-based investigation of brain structure in male adolescents with autistic spectrum disorder. *Neuroimage*, *22*(2), 619–625.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children (WISC III-UK)*. New York: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS)* (3rd UK ed.). London: Psychological Corporation.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, *18*(6), 469–474.

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., et al. (2006). The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Developmental Science*, *9*(3), 237–255; discussion 265–269.

Appendix

Strange Stories

The complete sets of stories are included as follows. The mental state stories and the unlinked sentences are suitable for use with both adults and children; the new questions for the unlinked sentences are a more suitable control task. The original physical state stories are most suitable for high-functioning adults, whereas the four new physical stories in addition to the first four (as listed) original physical stories are more suitable for children. The animal and nature sets are matched for difficulty to the other sets in children.

Mental State Stories

Simon is a big liar. Simon's brother Jim knows this, he knows that Simon never tells the truth! Now yesterday Simon stole Jim's ping-pong paddle, and Jim knows Simon has hidden it somewhere, though he can't find it. He's very cross. So he finds Simon and he says, ''Where is my ping-pong paddle? You must have hidden it either in the cupboard or under your bed, because I've looked everywhere else. Where is it, in the cupboard or under your bed''? Simon tells him the paddle is under his bed.

Q: Why will Jim look in the cupboard for the paddle?

2 points—reference to Jim knowing Simon lies

1 point—reference to facts (that's where it really is, Simon's a big liar) or Simon hiding it without reference to implications of lying

0 points—reference to general nonspecific information (because he looked everywhere else)

During the war, the Red army captures a member of the Blue army. They want him to tell them where

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

Q: Why did the prisoner say that?

2 points—reference to fact that other army will not believe and hence look in other place, reference to prisoner's realization that that's what they'll do, or reference to double bluff

1 point—reference to outcome (to save his army's tanks) or to mislead them

0 points—reference to motivation that misses the point of double bluff (he was scared)

Brian is always hungry. Today at school it is his favourite meal—sausages and beans. He is a very greedy boy, and he would like to have more sausages than anybody else, even though his mother will have made him a lovely meal when he gets home! But everyone is allowed two sausages and no more. When it is Brian's turn to be served, he says, ''Oh, please can I have four sausages, because I won't be having any dinner when I get home!''

Q: Why does Brian say this?

2 points—reference to fact that he's trying to elicit sympathy, being deceptive

1 point—reference to his state (greedy), outcome (to get more sausages) or factual

0 points—reference to a motivation that misses the point of sympathy elicitation/deception, or factually incorrect

Jill wanted to buy a kitten, so she went to see Mrs. Smith, who had lots of kittens she didn't want. Now Mrs. Smith loved the kittens, and she wouldn't do anything to harm them, though she couldn't keep them all herself. When Jill visited she wasn't sure she wanted one of Mrs. Smith's kittens, since they were all males and she had wanted a female. But Mrs. Smith said, ''If no one buys the kittens I'll just have to drown them!''

Q: Why did Mrs. Smith say that?

2 points—reference to persuasion, manipulating feelings, trying to induce guilt/pity

1 point—reference to outcome (to sell them or get rid of them in a way which implies not drowning) or simple motivation (to make Jill sad)

0 points—reference to general knowledge or dilemma without realization that the statement was not true (she's a horrible woman)

One day Aunt Jane came to visit Peter. Now Peter loves his aunt very much, but today she is wearing a new hat; a new hat which Peter thinks is very ugly indeed. Peter thinks his aunt looks silly in it, and much nicer in her old hat. But when Aunt Jane asks Peter, ''How do you like my new hat?,'' Peter says, ''Oh, its very nice.''

Q: Why does he say that?

2 points—reference to white lie or wanting to spare her feelings; some implication that this is for aunt's benefit rather than just for his, desire to avoid rudeness or insult

1 point—reference to trait (he's a nice boy) or relationship (he likes his aunt); purely motivational (so she won't shout at him) with no reference to aunt's thoughts or feelings; incomplete explanation (he's lying, he's pretending).

0 points—reference to irrelevant or incorrect facts/feelings (he likes the hat, he wants to trick her)

Helen waited all year for Christmas, because she knew at Christmas she could ask her parents for a rabbit. Helen wanted a rabbit more than anything in the world. At last Christmas Day arrived, and Helen ran to unwrap the big box her parents had given her. She felt sure it would contain a little rabbit in a cage. But when she opened it, with all the family standing round, she found her present was just a boring old set of encyclopedias, which Helen did not want at all! Still, when Helen's parents asked her how she liked her Christmas present, she said, ''It's lovely, thank you. It's just what I wanted.''

Q: Why did she say this?

2 points—reference to white lie or wanting to spare their feelings; some implication that this is for parent's benefit rather than just for her, desire to avoid rudeness or insult

1 point—reference to trait (she's a nice girl) or relationship (she likes her parents); purely motivational (so they won't shout at her) with no reference to parent's thoughts or feelings; incomplete explanation (she's lying, she's pretending)

0 points—reference to irrelevant or incorrect facts/feelings (she likes the present, she wants to trick them)

Late one night old Mrs. Peabody is walking home. She doesn't like walking home alone in the dark because she is always afraid that someone will attack her and rob her. She really is a very nervous person! Suddenly, out of the shadows comes a man. He wants to ask Mrs. Peabody what time it is, so he walks toward her. When Mrs. Peabody sees the man

coming toward her, she starts to tremble and says, ''Take my purse, just don't hurt me please!''

Q: Why did she say that?

2 points—reference to her belief that he was going to mug her or her ignorance of his real intention

1 point—reference to her trait (she's nervous) or state (she's scared) or intention (so he wouldn't hurt her) without suggestion that fear was unnecessary

0 points—factually incorrect/irrelevant answers; reference to the man actually intending to attack her

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

Q: Why did the burglar do that?

2 points—reference to belief that policeman knew that he'd burgled the shop

1 point—reference to something factually correct in story

0 points—factually incorrect/irrelevant answers

Original Physical Stories

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

Q: Why did the Blue army win?

2 points—reference to both weather conditions and either relative ground superiority or inability of other army's planes to be useful in fog (names of armies unimportant)

1 point—reference either to weather or relative superiority on ground versus air (because it was foggy); nothing about why weather makes it especially difficult for planes or nothing about planes being affected more than tanks; reference to fog to justify incorrect response (the aeroplanes won because the fog meant they could hide from the tanks)

0 points—reference to irrelevant or incorrect information (they won because they had better planes); justifications for why tanks are better than planes

A burglar is about to break into a jewelers' shop. He skillfully picks the lock on the shop door. Carefully he steps over the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the store-room and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, toward the shop door. Immediately the alarm sounds.

Q: Why did the alarm go off?

2 points—reference to animal which the burglar disturbed setting off alarm by crossing beam (type of animal unimportant)

1 point—reference to burglar setting off alarm (he was startled by the animal so crossed the beam); reference to animal setting off alarm without explaining it crossed the beam (he trod on a cat and it set off the alarm)

0 points—reference to irrelevant or incorrect factors (the animal's screech set off the alarm); alternative reasons for alarm going off (a security camera saw him and set the alarm off)

[A slight change has been made to this story from the original, which read ''Carefully he crawls under . . . ,'' as many participants in Experiment 1 found the wording confusing.]

Old Mrs. Robinson is very frail. One day she slips on her icy door step and falls on her side. She gets up right away, although she feels quite bruised and shaken. The next day her leg feels very stiff and she can scarcely walk. She makes her way to the doctors. As soon as the doctor hears about the fall, and sees her swollen side, he says, ''Go immediately to the hospital.'' At the hospital they take an X-ray.

Q: Why did they take an X-ray?

2 points—reference to possibility that she has fractured/broken her hip/leg; reference to wanting to know or trying to find out (i.e., ''it was broken'' is not enough); must refer to fact that X-rays are for broken things or bones (to see if there's any damage to the bone)

1 point—reference to general aim (to see what's wrong, because of her fall she might have damaged something) or factually correct (it's bruised and stiff)

0 points—reference to irrelevant (because she fell) or incorrect factors (that's what doctors do) or to X-rays being cures themselves (to mend her leg)

John is going shopping. He buys a nice new desk lamp, for his study. He needs a light bulb for his new lamp. He goes from the furniture department to the electrical department. In the electrical department he finds that there are two brands of light bulb of the right kind. Everbrite light bulbs cost less in single packs than Literite bulbs. However, only Literite bulbs come in multipacks of six. John buys the multipack, even though he only needs one bulb.

Q: Why does John buy the Literite bulbs?

2 points—reference to saving money by buying the multipack

1 point—reference to convenience of having more bulbs, or future need for more than one bulb; no mention of saving money

0 points—reference to irrelevant or incorrect factors (Literite bulbs are brighter)

Mrs. Simpson, the librarian, receives a special book which she has to catalogue and find an appropriate place for. She has to decide which section to file it under. The library is very big, and has different sections on many different subjects. The new book is about plants and their medical uses, and is heavily illustrated. However, Mrs. Simpson does not put it on the shelf with the rest of the books on botany. Neither does she put it with the books on medicine. Instead, she carefully takes it into a separate room. In this room all the books are kept in special cases, and the temperature is kept constant.

Q: Why did she do this?

2 points—reference to avoiding damage to the book because it is special

1 point—reference to the fact that the book is special; no reference to why it might be kept in a special case

0 points—reference to irrelevant or incorrect factors (she doesn't know where else to put it)

Henry is preparing for a big dinner party. He is famous for his excellent mayonnaise. He has bought lots of fresh eggs. The recipe says, ''Carefully separate the yolks of six eggs and add oil very gradually.'' He has already bought easily enough dessert to feed everyone. However, he now looks up the recipe for meringues. Henry will not waste anything.

Q: Why does Henry make meringues?

2 points—reference to Henry not liking to waste anything and therefore using up the left-over egg whites

1 point—reference either to not wasting anything or to having left-over egg whites

0 points—reference to irrelevant or incorrect factors (he's having a party)

Paul is very rich, and today he is going to buy an expensive new car. He is considering whether to make a single payment, or whether to spread the cost over the year. If he pays in monthly installments, the dealer will charge 5% interest on the loan. His bank currently gives him 8% interest on the money in his account. Even though he has easily enough money to pay the full amount, he decides to pay by monthly installments.

Q: Why does he do that?

2 points—reference to getting more interest from the bank than he'd pay on the loan and therefore to saving money

1 point—reference to saving money; no explanation why he'd save money

0 points—reference to irrelevant or incorrect factors (he doesn't have enough money)

Sarah is very far-sighted. She has only one pair of glasses, which she keeps losing. Today she has lost her glasses again and she needs to find them. She had them yesterday evening when she looked up the television programs. She must have left them somewhere that she has been today. She asks Ted to find her glasses. She tells him that today she went to her regular early morning exercise class, then to the post office, and last to the flower shop. Ted goes straight to the post office.

Q: Why is the post office the most likely place to look?

2 points—reference to post office being place she would most likely use her glasses (to read/write/look at stamps etc); may talk about either putting glasses on or taking them off

1 point—plausible alternative reason for being in post office (there are lots of people there, you might have posted them by mistake, people take lost things there)

0 points—reference to irrelevant or incorrect factors (that was the last place she went, you can buy glasses at the post office, she needed the glasses to hear better); general factors, nonspecific to post offices

New Human Physical Stories

Bob and Jim are best friends. They are both 10 years old. Bob has brown hair, green eyes and is over 5 feet tall. Jim looks very different to Bob. He

has blonde hair and blue eyes and he is much smaller than Bob. Bob and Jim go on an outing to the fun fair. They go on lots of rides. For the last ride of the day they decide to go on the big rollercoaster. But there is a sign which says: For safety reasons no persons under 5 feet are allowed on.

Q: Why does only Bob go on the rollercoaster?

2 points—reference to Jim being too short for the ride or Bob being tall enough (Jim's less than 5 feet)

1 point—reference to Jim being short or Bob being tall or both; no reference to height in comparison to the limit (Jim's shorter than Bob)

0 points—reference to irrelevant or incorrect factors (Jim doesn't like rollercoasters)

Rupert has never been skiing before and is looking forward to his first skiing holiday this winter. All his kit for the holiday has been well prepared; his mum has bought him a pair of goggles and she has thoroughly waxed and polished the bottom of his skis to protect them. On the first day of Rupert's holiday his skis keep slipping from underneath him, making him fall over into the snow.

Q: Why does Rupert keep falling over?

2 points—reference to Rupert's Mum having waxed the skis, making them slippery

1 point—reference to Rupert's never having skied before

0 points—reference to irrelevant or incorrect factors (his skis are loose)

Clare is having her room redecorated; her mother is painting the walls and having new curtains hung. Before, Clare's room was pink and white with thin net curtains but now the walls are dark red, and brand new thick and expensive velvet curtains have been put up. On the first morning in her new room, Clare fails to wake up at the normal time. As her mother rushes to get her out of bed for school, Clare says it must be too early to get up because it ''feels like the middle of the night.''

Q: Why did Clare oversleep?

2 points—reference to the room being darker after redecoration (her room is dark now that she has thicker curtains)

1 point—reference to redecoration; no reference to this making the room darker

0 points—reference to irrelevant or incorrect factors (she's too tired, she doesn't want to go to school)

Sam decides to go on a long walk to get some fresh air. Unfortunately, just after leaving the house, the wind begins to pick up and it starts to rain. Luckily Sam always has an umbrella with him. He quickly puts up the umbrella and wraps his coat tightly around him. Suddenly a gust of wind blows the umbrella straight out of Sam's hand and it lands in a large, very prickly bush. Sam manages to run and fetch it before it blows off again and is pleased to find it all in one piece. As he walks home, he notices that his head is starting to get wet despite the umbrella.

Q: Why is Sam getting wet?

2 points—reference to the bush making holes in the umbrella

1 point—reference to either the bush or to holes in the umbrella

0 points—reference to irrelevant or incorrect factors (it was raining, he hasn't got an umbrella)

Animal Stories

Emperor penguins live in the Antarctica, where it is extremely cold. There is always snow on the ground and ice on the surface of the sea. Emperor penguins can often be found standing clumped together in huge, huddled masses. Every few minutes, a penguin in the middle of the huddle moves to the edge of the huddle, changing places with one of the penguins on the outside of the group.

Q: Why do the penguins keep changing places?

2 points—reference to the middle penguin being the warmest or to taking turns at being warm

1 point—reference to keeping warm without relating this to the huddle

0 points—reference to irrelevant or incorrect factors

Snakes are remarkable animals. They have very stretchy skin, which they shed once a year and can also separate their upper and lower jaws and open their mouths really wide. The anaconda is an example of a very large snake. One day, a deer gallops under a tree, from which an anaconda is hanging, as it makes its way toward a lake. Later that day, the snake is lying on the ground with a huge bulge in its middle. The deer however is nowhere to be seen.

Q: Where is the deer?

2 points—reference to the snake having eaten the deer (in the snake's tummy)

1 point—reference to the deer having been eaten or being dead without reference to the snake

0 points—reference to irrelevant or incorrect factors

It has been raining for days and days and there are no signs that it is going to stop anytime soon. A

little island lies in the middle of a huge river. The water in the river has been slowly rising each day and it has nearly reached the top of the river banks. The otters swim around in the water and the field mice run about the island gathering food. Five days later, the rain has finally stopped. The otters still swim in the water, but there are no signs of the field mice.

Q: What has happened to the field mice?

2 points—reference to the field mice having drown or being dead because of the water

1 point—reference to them being dead without reference to water or to the water taking them away without reference to them being dead

0 points—reference to irrelevant or incorrect factors

Lions are fierce hunters. They can run as fast as a car when they are young and fit but they get very slow and weak when they are old. One very hot day, an old and hungry lion is standing at the mouth of a cave, watching a herd of zebras moving across a large open plain. When the herd has passed by, the lion begins to chase a small zebra at the back of the herd. One by one, the zebras nimbly jump across a river. But the lion returns to the cave, still hungry.

Q: Why is the lion still hungry?

2 points—reference to the lion being old or slow or weak and so not being able to catch the zebras (he was too old and weak to jump over the river)

1 point—reference to the lion being unable to catch the zebras without reference to him being old or slow or weak (he couldn't jump across the river to catch the zebras)

0 points—reference to irrelevant or incorrect factors

Some types of birds, like geese and swallows, only like very warm weather. When it is winter in England, it is still very warm in other countries that are further south. Last autumn, flocks of swallows could be seen flying in huge groups in the same direction away from England. At the beginning of summer, these swallows flew back to England.

Q: What were the swallows doing?

2 points—reference to migration or to flying to the place where it is currently warm

1 point—reference to specific examples without a general understanding of finding warm places (they were going away from the cold weather in England; they were going away to a hotter country) or to keeping warm without any explanation

0 points—reference to irrelevant or incorrect factors

Swordfish come in many different colors; black, grayish blue, brown, purple, and bronze. They often live in tropical places where the water is very warm. In these tropical places, the weather sometimes gets so hot that thunderstorms occur. When this happens, huge waves crash onto the beaches and travel a long way up the land. On one very hot day in Hawaii, a swordfish is lying on the beach.

Q: Why is the swordfish lying there?

2 points—reference to the swordfish having been washed up onto the beach by a huge wave or the storm

1 point—reference to the storm or sea or waves without reference to the swordfish

0 points—reference to irrelevant or incorrect factors

Seals have very big eyes and long whiskers that help them to sense tiny movements. Underwater they use their whiskers to find fish so they can then catch the fish and eat them. A seal without any whiskers at all, is lying on a rock in the North Sea. This seal is very, very thin and tired.

Q: Why is this seal so thin?

2 points—reference to the seal having no whiskers and therefore not being able to find fish

1 point—reference either to not having eaten or to not having whiskers

0 points—reference to irrelevant or incorrect factors

Animals that live in groups often have an order of importance within the group. The strongest male is the leader of the group. This leader will often attack other animals in the group who are not as strong as this leader. This shows the other animals how important the leader is. In a group of chickens a very small chicken hasn't got many of its feathers left.

Q: Why hasn't this chicken got many feathers left?

2 points—reference to the chicken having been attacked by the leader/a larger chicken or having been attacked because of its size

1 point—reference either to the chicken's size or to having been attacked (reference to it being young rather than small do not count)

0 points—reference to irrelevant or incorrect factors

Natural Physical State Stories

In stormy weather, rocks often fall from the top of mountains. One day on a mountain in the Dolomites, a very large boulder becomes loose and

starts rolling down the mountain. It rolls and rolls and rolls, gathering speed and spinning and bouncing off the mountain side. Suddenly, there is a very noisy splash.

Q: Why is there a loud splash?

2 points—reference to the boulder falling into water to make the splash (the boulder must have fallen into a lake)

1 point—reference to water without reference to the boulder (there was a pool at the bottom of the mountain)

0 points—reference to irrelevant or incorrect factors (it's very big so it's very noisy)

A storm is building up over a little village in the mountains. There is thunder and lightening. The trees sway in the heavy gusts of wind, and the rain is pouring down. Leaves and even some branches are falling from the trees. After one extremely bright flash of lightning, there is a loud crashing noise and the lights go out in all of the houses in the village.

Q: Why did the lights go out?

2 points—reference to the lightning hitting a tree which fell onto a power line and cut the electricity (the lightning hit a tree which crashed into the electricity wires)

1 point—reference to lightning hitting power lines

0 points—reference to irrelevant or incorrect factors

It is a very cold winter and has been snowing for days and days. The snow has covered everything; the trees, the houses, the hilltops, even the fences are covered in a thick layer of snow. Everything looks completely white apart from the dull gray sky. One morning, the skies are blue and the sun comes out. The sun beats down on the houses, the trees, the hilltops, and the fences. Puddles start to form at the edges of the fields.

Q: Why are there lots of puddles?

2 points—reference to the snow melting or the effect of the sun on the snow (the sun makes the snow turn into water)

1 point—reference to the snow without reference to melting or the sun (because of the snow)

0 points—reference to irrelevant or incorrect factors

The little village of Littlehurst is close to the river Worrow. A year ago, a wall was built all the way round the edge of the village. The river floods its banks in April every year and, in the past, water would flow into many houses and cause lots of damage. For 3 weeks now the rain has been pouring down. However, this year, all the houses in Littlehurst are perfectly dry inside.

Q: Why were all the houses dry?

2 points—reference to the wall stopping the water entering the village

1 point—reference either to the wall or to the water not getting in but no connection between the two

0 points—reference to irrelevant or incorrect factors

The summer has been long and very warm, just the right conditions for producing lots of apples. All summer long the orchard has been quiet and peaceful. Now, at the end of summer, the apples hang from the trees, glistening in the bright sun, all ripe and rosy. And every now and then in the orchard, little thumps can be heard.

Q: Why are there little thumps?

2 points—reference to the apples falling from the trees or hitting the ground

1 point—reference to the apples without mention of them falling or hitting the ground

0 points—reference to irrelevant or incorrect factors

One of the huge parks in the middle of London has stone statues all around the edge of it. It also has lots of trees, which drop their leaves in autumn every year. On a cold, dry morning in November, a huge bonfire is burning all the leaves in one corner of the park and the statues in that corner of the park can't be seen. But in the afternoon, it is clear that they are still there.

Q: What had happened in the morning?

2 points—reference to the smoke from the fire covering the statues

1 point—reference to the bonfire or smoke without explaining how this affected the statues

0 points—reference to irrelevant or incorrect factors including incorrect reference to the fire (it burnt down the statues)

It is late in April and the sky has been clear and blue all morning. An empty fountain sparkles in the middle of the park. As the day goes on, the sun occasionally disappears behind little white fluffy clouds, soon to appear again on the other side. However in the late afternoon, the sky becomes dark and filled with lots of gray clouds. A little after this, the fountain starts to spout water.

Q: Why did the fountain spout water?

2 points—reference to the rain filling up the empty fountain

1 point—reference to either to rain or to the fountain being filled up
0 points—reference to irrelevant or incorrect factors

Iceland is a country where earthquakes often occur. They happen very suddenly when big rocks under the ground suddenly move, making the ground shake. One day last year, the ground started to shake near a mountain in the south of Iceland. As the ground shook more and more, a large cloud of smoke appeared above the mountain and huge flames roared from the mountain top.
Q: Why did this happen?
2 points—reference to a volcano or to an earthquake starting a volcano (reference to larva or similar is also fine)
1 point—reference to the earthquake (because the ground under mountain was shaking)
0 points—reference to irrelevant or incorrect factors

Unlinked Sentences (First Question Is Original Question; Second Is New Question)

The two countries had been at war. A housewife is about to enter the super-market. Today he is going to buy an expensive new stereo. Mrs. Brown, the postmistress, receives a special parcel. Mrs. Pearson wouldn't harm a fly. Mary's birthday is in February. Late one evening the old man was watching television.
Q: Is Mary's birthday in February?
2 points—Yes
0 points—No
Q: When is Mary's birthday?
2 points—February
0 points—Anything else

Young Simon is very robust. She sees that Fred cannot play. Jeremy is always laughing. Ruth sees her uncle very often, but today he has gone to Brazil. Richard is packing up to go away. Today, at college, it is Jim's worst lecture—statistical mechanics. She has only one dollar left, which she must keep for her bus fare. He buys a bright tie, to go with his new shirt.
Q: Does she have $2 left?
2 points—No
0 points—Yes
Q: How many dollars does she have left?
2 points—One
0 points—Anything else

Simon takes the special butter from the refrigerator. Each boxer has won several fights. He skillfully picks out the imperfect items. They are either in Boston or in New York. She has to cut the grass and find somewhere to plant the bay tree. The conductor sees that the cellist has broken a string. Tracy took the bus to the station.
Q: Did Tracy take the bus?
2 points—Yes
0 points—No
Q: Where did Tracy take the bus to?
2 points—The station
1 point—Anything that seems part way there (e.g., the bus station)
0 points—Anything else

The four brothers stood aside to make room for their sister, Stella. Gill repeated the experiment, several times. The name of the airport has changed. Louise uncorked a little bottle of oil. The 2 children had to abandon their daily walk. She took a suite in a grand hotel. It was already 20 years since the operation.
Q: Did the children take their walk?
2 points—No
0 points—Yes
Q: Who abandoned their daily walk?
2 points—The 2 children
1 point—Anything that seems part way there (e.g., the children)
0 points—Anything else

One day Uncle Simon came to visit Alex. The first part of the performance had come to an end. He put away the letter and stuck his hands in his pockets. She was still holding her umbrella. The cats ran back to the boy. Flora came into the middle of the square. The little island had a high rocky shoreline.
Q: Did Flora go to the square?
2 points—Yes
0 points—No
Q: Where did Flora go to?
2 points—The middle of the square
1 point—Anything that seems part way there (e.g., the square)
0 points—Anything else

At the edge of the road a little grass was growing. He reaches out to find the light switch. A sailor who has just left his ship is walking to the town. She has to decide where to keep the pasta. At last daylight came, and Tommy got out of bed to open his presents. Jim knows all about investing money,

as he works in a large bank. They exchanged a few brief words about the weather.

Q: Did Tommy stay in bed?

2 points—No

0 points—Yes

Q: Why did Tommy get out of bed?

2 points—To open his presents

1 point—Anything that seems part way there (e.g., coz daylight came/it was morning)

0 points—Anything else

She is always saying that someone will eventually find the treasure. Everyone is allowed two visits and no more. At the psychiatry department they were interviewing the new nurses. Jim will win the first race of the meeting. She has taken all the children to visit the zoo today. Simon's uncle is wearing a new suit. The same phrase of twenty-three notes recurred throughout.

Q: Will Jim lose the first race?

2 points—No

0 points—Yes

Q: What will Jim win?

2 points—The first race of the meeting/the first race

1 point—Anything that seems part way there (e.g., the race)

0 points—Anything else

He needs a new engine for his old car. The prize is an immediate lump sum of $20,000 tax-free. Japan is stronger than Italy in economic terms. The mother is very brave and long suffering. The new book is about statistics and experimental design, and contains many graphs. The front room contained a little bird in a cage. Although Jim is only 29 years old, he has an income of $20,000 per year. There are not many people this evening in the large rectangular dining room.

Q: Is the mother brave?

2 points—Yes

0 points—No

Q: Who is brave and long suffering?

2 points—The mother

0 points—Anything else