Journal of Child Psychology and Psychiatry 43:4 (2002), pp 517–528

A new 'advanced' test of theory of mind: evidence from children and adolescents with Asperger syndrome

Nils Kaland,[1] Annette Møller-Nielsen,[2] Kirsten Callesen,[2] Erik Lykke Mortensen,[3] Dorte Gottlieb,[4] and Lars Smith[5]

[1]Lillehammer College, Norway; [2]The Danish University of Education, Denmark; [3]University of Copenhagen, Denmark; [4]The Psychological Unit of Copenhagen County, Denmark; [5]University of Oslo, Norway

Background: The aim of the present study was to assess the ability of children and adolescents with Asperger syndrome (AS) ($N=21$) to infer *physical* versus *mental* states on a new 'advanced' test of theory-of-mind – *Stories from Everyday Life*. The participants in the AS group were of normal intelligence and were compared with an age-matched control group ($N=20$) of normally developing children and adolescents. **Method:** The test materials comprised 26 short stories or 13 pairs of different types of stories. This contextually complex theory-of-mind battery aimed to record the participants' ability to make inferences about *physical* as well as *mental* states. The first part of each story described a *physical* or *mechanical* event, and a test question then tested the participants' ability to make an inference about a *physical state*. The last part of the stories contained two questions, testing the ability to infer a *mental state* from the story context, e.g., understanding social communication such as a *lie*, *white lie*, *figure of speech*, *misunderstanding*, *double bluff*, *irony*, *persuasion*, *contrary emotions*, *forgetting*, *jealousy*, *intentions*, *empathy* and *social blunders*. The participants' reaction time and number of prompt questions were also recorded. **Results:** The participants in the AS group showed significantly more problems attributing *mental state* inferences relative to the control group. They performed considerably better on tasks involving a *physical state*, but still did less well than the controls. They had a tendency to interpret behaviour and utterances literally, without regard to context, and to choose a *physical explanation* when a *mental state* answer was more appropriate. They also needed significantly more prompt questions and used significantly more time than the controls to solve the tasks, especially those involving *mental state* inference. **Conclusions:** This study supports earlier findings that individuals with AS/HFA have difficulties attributing *mental states* in context, but seem to have significantly fewer difficulties inferring *physical states*. The fact that the clinical group also used significant longer reaction time and needed significantly more prompt questions to solve the tasks relative to the control group may also be related to their problems in understanding *mental states*. However, the possibility remains that these difficulties could represent a separate factor – or a distinct 'cognitive style' – suggesting that at least some individuals with AS may be generally slow in solving cognitive tasks. **Keywords:** Asperger syndrome, cognitive tests, theory of mind, 'naturalistic' test battery, mental vs. physical inference, reaction time. **Abbreviations:** AS = Asperger syndrome.

Asperger syndrome (AS) is a type of pervasive development disorder (PDD) recently included in the ICD-10 (WHO, 1993) and the DSM-IV (APA, 1994). The syndrome is characterised by deficits in reciprocal social interaction of the autistic kind, subtle impairment of communication and the presence of idiosyncratic isolated interests. According to ICD-10 and DSM-IV there is no history of language delay, and intelligence is in the normal range. The condition has been termed Asperger syndrome (Asperger, 1944) after the Austrian physician who observed a number of cases with clinical features resembling Kanner's (1943) description of autism. However, Asperger's description differed from Kanner's in that speech was less commonly delayed, motor deficits and clumsiness (Ghaziuddin & Butler, 1998) more common, the onset of the disorder somewhat delayed, and all the initial cases occurred in boys (Wing, 1981). Asperger also suggested that similar problems could be observed in family members, particularly among fathers. However, the validity of AS as distinct from other conditions, notably the other pervasive developmental disorders, remains controversial (Klin, Volkmar, Sparrow, Cicchetti, & Rourke 1995; Volkmar, Klin, Schultz, Rubin, & Bronen, 2000; Wing, 1991).

Previous studies have found a sizeable subgroup of children and adults with AS or high-functioning autism (Bauminger & Kasari, 1999; Bowler, 1992; Dahlgren & Trillingsgaard, 1996; Happé, 1994; Ozonoff & McEvoy, 1994; Ozonoff, Rogers, & Pennington, 1991) who pass second-order tests of theory of mind, although they share the social and communicative symptoms of autism. Second-order theory-of-mind tests demand that the participant reasons about what one person thinks about another person's thoughts (Perner & Wimmer, 1985; Baron-Cohen, 1989). Tager-Flusberg and Sullivan

© Association for Child Psychology and Psychiatry, 2002.
Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA

(1994) reported that children with autism who passed first-order theory-of-mind tasks also could pass second-order tasks if the information processing demands were minimised. Clearly, these studies cannot be taken as conclusive evidence of an intact theory of mind in individuals with autism or AS. Such second-order tests have a ceiling corresponding to a mental age of about 6 years or less. It is therefore impossible to determine whether autistic individuals (typically much older than age 6) are intact or impaired in their theory-of-mind skills.

Happé (1994), taking this issue seriously, tested able individuals with autism on an 'advanced' theory-of-mind task. This involved story comprehension, where the key questions concerned either a character's mental state or physical events. She found that even competent persons with autism had more difficulties with the mental state stories than did matched controls, and that they used fewer and more inappropriate mental state terms than the control group in their justifications of why the characters behaved as they did. Jolliffe and Baron-Cohen (1999) later replicated this finding using a modified version of the story battery with individuals with AS.

More recently, other types of theory-of-mind tasks have demonstrated deficits in high-functioning individuals with autism or AS. Abell, Happé, and Frith (2000) and Klin (2000) found deficits in the attribution of mental states to animated shapes in high-functioning individuals with autism. Heavey, Phillips, Baron-Cohen, and Rutter (2000), using excerpts of films showing characters – in social situations – developed to approximate the demands of real-life mentalising, found that adults with high-functioning autism and AS were impaired in their ability to answer the questions requiring mind-reading abilities. Baron-Cohen, Jolliffe, Mortimore, and Robertson (1997), and Baron-Cohen, Wheelright, Hill, Raste, and Plumb (2001), using a 'pure' theory-of-mind task (the *Eyes Task*) involving inference of the mental state of a person just from the information conveyed in photographs of his or her eyes, found that adult individuals with autism and AS were significantly impaired on this task – relative to age-matched normal controls. Kaland (2000) replicated these results with children and adolescents with AS, using the revised version of the task. Kleinman, Marciano, and Ault (2001) found that individuals with AS had difficulties in their determination of voice intonation, suggesting that other modalities than the eyes account for these individuals' difficulties in attributing mental states.

The aim of the current study was to investigate whether children and adolescents with AS would be able to infer mental state concepts when the theory-of-mind elements were embedded in a 'naturalistic' story context. Another important aim was to examine whether these persons have more success with physical than with mental state inferences in the context of story comprehension. We tested theory-of-mind performance and non-theory-of-mind performance, respectively, on a new test battery – *Stories from Everyday Life*. This test battery is essentially designed after Happé's *Strange stories* (1994), thus extending her line of research. However, it was designed to be contextually somewhat more complex than hers. The tasks involve making inferences about physical and mental states from a set of vignettes, or stories, about everyday situations. All the 26 stories were designed to probe the ability to infer physical as well as mental states in a story context. The *physical state* components were embedded in the first part of the story text, often in the first or second paragraph. The *mental state* components were embedded mainly in the last part of the story text (see method and Appendix 1 below).

It was hypothesised that the participants with AS would have fewer difficulties inferring physical than mental states in context, and even perform as well, or nearly as well, as the control group on the physical tasks. It was also expected that the control group was not likely to have problems with either kind of task.

As yet, relatively little attention has been paid to the reaction times of persons with AS when performing theory-of-mind tasks (Bowler, 1997; Carruthers, 1996). Bowler (1997) found that high-functioning individuals with autism used more time than controls in answering the mental state questions of standard false belief tasks, and that they were also slower than the controls on the physical state control questions. We therefore recorded the reaction time on set B of the *Stories from Everyday Life*. It was expected that the participants with AS, as a group, would take more time than the controls to answer the test items, especially the ones involving inferences about mental states, and perhaps also those concerning physical states. It was also expected that the participants in the AS group would need significantly more prompt questions than the controls in solving the tasks.

The participants in the clinical group were given first- and second-order standard theory-of-mind tests. The first-order task was Perner, Frith, Leslie, and Leekam's (1989) *Smarties task*. The second order task was a version of Tager-Flusberg and Sullivan's (1994) *Birthday present task*. All participants but one (95.2%) passed each of these tasks. Participants were also tested on Happé's *Strange stories* task (1994), another advanced theory-of-mind test, consisting of 24 short vignettes, with 12 types of story and 2 examples of each. Examples of the 12 story types are *Double Bluff*, *Figure of Speech*, *Joke*, *Lie*, *Misunderstanding*, *Persuade*, *Pretend*, *Sarcasm* and *White Lie*. The test contains two test questions, a *comprehension* question and a *justification* question.

Method

Participants

Twenty-one children and adolescents with Asperger syndrome and with verbal intelligence within the normal range and 20 normal controls participated. The participants in the AS group were recruited from the Kasperskolen in Copenhagen – a special school for children and adolescents with Pervasive Developmental Disorders (PDD), and from the Centre for Autism in Copenhagen. They had been diagnosed in clinical settings – according to the ICD-10 criteria for Asperger syndrome – by experienced child psychiatrists, who had no knowledge of the purpose or the design of the study. Four persons with PDD, who did not fully fulfil the ICD-10-criteria for Asperger syndrome, were excluded from the study before the test period started. Thus the selection of participants was as stringent as possible, given the nosologically controversial diagnosis of AS. Males are over-represented as far as AS is concerned (Ehlers & Gillberg, 1993) and, since females probably have a somewhat different clinical symptomatology than males (Attwood, 1998; Gillberg, 1997), only males were included in the current study.

Controls were volunteers recruited from different schools in Copenhagen. At the time of the study they had no known history of developmental lag or disorder. The two groups were matched on age, gender and parental socio-economic status (SES). Since the control group scored significantly higher than the AS group on verbal IQ, the importance of this variable was controlled by covariance analysis. Participant characteristics are shown in Table 1.

The statistical analysis showed no significant age difference between the two groups ($df = 1.39$; $F = .08$, $p < .8$). All participants had verbal IQ above 90 and full-scale IQ above 81 when tested on the WISC-III (Wechsler, 1998). The control group performed significantly better than the clinical group on verbal IQ ($df = 1.39$, $F = 11.48$, $p < .001$), performance intelligence (PIQ) ($df = 1.39$, $F = 10.64$, $p < .002$), and on full-scale intelligence (FIQ) ($df = 1.39$, $F = 13.25$, $p < .001$).

Analysis of questionnaires, filled out by the parents, revealed no significant group differences in SES, level of parental education ($df = 1.38$, $F = .02$, $p = .90$), and parental occupational experience ($df = 1.38$, $F = 1.22$, $p = .28$).

Two of the participants in the AS group were left-handed as opposed to four in the control group. All participants were free of medication at the time of testing.

Story material

The test material comprised 26 short stories or 13 pairs of different types of stories. The stories were split into an A and a B set and included *lie*, *white lie*, *figure of speech*, *misunderstanding*, *double bluff*, *irony*, *persuasion*, *contrary emotions*, *forgetting*, *jealousy*, *intentions*, *empathy* and *social blunders*. We found it sufficient to record the reaction time on only half of the stories, on the B set.

The action of each story proceeded towards a climax. The first part of each story described a physical or mechanical event, and a question tested the participants' ability to make an inference about a *physical state* when answering.

There were 10–15 questions to each story. Most of them were *control questions* asked to make sure that the participants had grasped the gist of the story. The aim of the central questions was to test the participants' ability to infer *physical* as well as *mental states* from the story context. One question, usually number 2–5, was asked in order to assess inference about a physical state, whereas the last two questions were asked in order to determine inference about a mental state. The two mentalistic test questions on the last part of the stories were the *comprehension question*, which usually took the form 'Is it true what X says?' and the *justification question*, which usually took the form 'Why does X say that?'

The following abstract from the second paragraph of the story type *irony* is an example of a text where the participant has to make an inference about a physical state: Two brothers, 8-year-old Tom and 14-year-old Adrian, are asked by their mother to tidy their rooms:

> Tom, the youngest of the brothers, is always making a mess, and his room is usually very untidy. His mother often complains about the mess. Adrian seldom has to hear such remarks, but his mother says that he should now and then help his father tidying the villa garden.

The *physical inference* question on the text was: 'How does Adrian's room look?' In order to answer this question the participants have to make an inference about a *physical state*.

The two boys go to their rooms to begin tidying. After a while their mother asks if they will soon be finished, and Adrian replies that he is finished. The last paragraph of the story is as follows:

> … eight-year-old Tom hasn't begun to tidy up at all! Adrian's mother asks if he can look in Tom's room to check if he has tidied up. Adrian opens the door to Tom's room, peers in, and sees that the room appears as it normally does. He shouts to his mother: 'Mother, Tom has as usual done a splendid job tidying up!'

Table 1 Characteristics of participants – *Stories from Everyday Life* (SEL). Group means (SD), age and range

Group	*N*	CA mean (*SD*)	Verbal IQ mean (*SD*)	Performance IQ mean (*SD*)	Full IQ mean (*SD*)
Asperger	21	15.72 (38.7)	111.40 (16.9)	98.30 (14.4)	106.40 (16.8)
Range		10:2–20:4	90–138	74–123	81–134
Control	20	15.58 (39.4)	126.40 (10.6)	113.10 (14.7)	122.70 (11.2)
Range		9:6–20:9	104–144	78–136	102–142

The *comprehension question* 'Is it true what he says?' was then asked. The next question, the *justification question*, was as follows: 'Why does Adrian say this?'

In order to be able to answer the justification question correctly, the participant has to make an inference about a mental state – *irony*; i.e., that Adrian is sarcastic. See Appendix 1 for a complete story example.

Procedure

Each participant was tested alone in a room that was free from distractions. The experimenter introduced each story as follows: 'Here are some stories and some questions. I am going to read the stories, and I'd like you to listen carefully and answer the questions at the end of each story. If you want, you may read the stories yourself while I am listening. OK? Let's start'. The experimenter sat next to the individual and read out each story. In some cases the participants preferred to read the story out loud to the experimenters. The sheet remained in front of the participant throughout the reading and questioning. This was done in order to minimise memory requirement.

A second experimenter recorded the answers that were provided. Breaks with refreshments were administered as needed, usually after 4–7 stories had been completed. The stories were printed on numbered pages without headings, and were presented in random order. Two versions of each story type were never presented one after the other. Positive comments were given throughout the test session in order to encourage the participants. No differential feedback was given for right versus wrong answers.

In most cases it was not necessary to ask all the control questions because the participants mostly provided complete answers, indicating understanding of the story content. Very few seemed to have striking difficulties with memory and *general* comprehension of the story text.

On the test questions, prompts were given if an incomplete answer was provided. A participant might give an answer to one of the test questions that was not quite correct. He was then given one or more additional questions in order to try to make it easier for him to understand the test question. So, for example, in one of the *persuasion stories* a rather aggressive car salesman tries to convince a customer to buy a second-hand car. On the test question the participants might answer that he was trying to sell the customer a new car. A prompt question, for example 'What is the salesman trying to do?' or 'What do we call what he is trying to do?', was then asked to see if the participant could give an even more precise answer. Asking him what the story character meant by saying/doing that, or simply asking the participant to read the last paragraph again, are examples of other types of prompts.

The physical information was always presented first in the story and the mental state information always towards the end of each story. In a few cases it was necessary for the participants to re-read the physical inference paragraph. To answer the *mental inference* questions the participants did not need to read the whole story again because the difficult components were largely embedded in the last paragraphs. If they provided insufficient answers to the mental-state inference question, they were sometimes encouraged to read the last, or last two, paragraphs again.

The younger participants in the control group received the stories in a way similar to the clinical group. The older ones mostly read the stories for themselves in the classroom and wrote down their answers on the sheet.

Reaction time

The time taken to answer the test questions was recorded (on the B set). The aim of this procedure was to test the prediction that the individuals with AS would be slower than the normally developing persons on some types of cognitive tasks, such as conceiving mental-state answers (cf., Bowler, 1997). A stopwatch was started as soon as the test question had been asked. If the answer given was wrong, the watch was stopped, and it was started again when a prompt question had been administered. The reaction time was recorded before and after a prompt question.

Scoring procedure

Since many of the justifications were only partly correct, it was decided to give 2 points for fully correct answers, 1 point for partly correct answers, and 0 for incorrect answers. A second rater was employed to validate this subjective judgement. Throughout the rating procedure, two raters were involved. However, none of them were blind to the diagnosis of the participants. The degree of inter-rater concordance was calculated for each story type. Mean values of kappa on the A stories set was .76 (.41–1.0) for *physical* inference and .85 (.69–1.0) for *mental* inference. On the B stories set the kappa values were .77 (.48–1.0) and .87 (.59–1.0) for *physical* and *mental* inference respectively (see Appendix 2). Values of kappa exceeding .8 represent substantial reliability, values between .4 and .8 represent fair to moderate agreement (Shrout, 1998).

Results

Table 2 lists the results on both sets of *Stories from Everyday Life*. As can be seen, the physical inference questions proved to be relatively easy for the AS participants. Their mean score was 23.9 (out of 26 possible), as opposed to the mean of the control group at 24.8. An analysis of variance (ANOVA) showed that the difference between the two groups was not statistically significant ($df = 1.39$, $F = 3.57$, $p = .07$). A covariance analysis (with control for age and VIQ) yielded a statistically significant difference between the two groups ($df = 1.37$, $F = 4.42$, $p < .05$).

The mental inference task, however, was considerably more difficult for the clinical group at a mean of 18.0 (out of 26 possible). In contrast, the control group earned a mean score of 24.6. ANOVA yielded a statistically significant difference between the groups ($df = 1.39$, $F = 50.91$, $p < .001$); controlling for age and VIQ still resulted in significantly better performance by the control group than the AS group: ($df = 1.37$, $F = 32.22$, $p < .001$).

Table 2 Physical inference (PI), mental inference (MI) performance, prompt questions (PQ) and reaction time (RT) on *Stories from Everyday Life* (SEL)

Group	Physical inference (PI)	Mental inference (MI)	Difference
A stories			
Inference:	mean (SD)	mean (SD)	mean (SD)
AS group	23.9 (1.7)	18.0 (3.9)	5.9 (3.9)
Control group	24.8 (1.4)	24.6 (1.5)	.2 (1.6)
Difference	−.9 sig	−6.6 sig	5.7 sig
Prompt questions:			
AS group	.8 (1.7)	13.6 (5.3)	12.8 (5.2)
Control group	.1 (.3)	2.0 (2.7)	1.9 (2.5)
	−.7	−11.6	10.8 sig
B stories			
Inference:	mean (SD)	mean (SD)	mean (SD)
AS group	22.1 (2.0)	17.8 (4.1)	4.3 (3.6)
Control group	24.2 (1.3)	24.5 (1.9)	.3 (1.3)
Difference	−2.1 sig	−6.7 sig	4.0 sig
Prompt questions:			
AS group	1.5 (1.4)	13.5 (5.3)	12.0 (5.1)
Control group	.3 (.7)	3.7 (3.3)	3.4 (3.3)
Difference	−1.2 sig	−9.8	8.6 sig
Reaction time, excl. prompts:			
AS group	56.5 (25.2)	98.2 (41.1)	41.7 (26.4)
Control group	20.6 (14.0)	37.1 (23.8)	16.5 (11.9)
Difference	−35.9 sig	−61.1 sig	25.2 sig
Reaction time, incl. Prompts:			
AS group	71.5 (39.8)	211.3 (144.0)	139.8 (116.3)
Control group	20.8 (14.1)	47.5 (36.0)	26.7 (24.1)
Difference	−50.7 sig	−163.8 sig	113.1 sig

On the B set of stories the AS persons scored a mean of 22.1 on the physical inference task, whereas the controls obtained a mean of 24.2. Interestingly, the respective mean values for the control group on the A and B stories were 24.8 and 24.2 for performance on *physical* inference and 24.6 and 24.5 for performance *on mental* inference.

An analysis of covariance demonstrated that the AS group did not perform at the level of the control group on *physical* inference ($df = 1.37$, $F = 13.41$, $p < .001$). As expected, however, the two groups performed significantly differently from each other on *mental* inference performance ($df = 1.37$, $F = 36.60$, $p < .001$).

On the A stories there was a statistically significant difference *within* the AS group between performance on *physical* and *mental* inference tasks (ANOVA, $df = 1.20$, $F = 47.99$, $p < .001$; Wilcoxon Test, $p < .001$). On the B stories there was a statistically significant difference between the same variables (ANOVA, $df = 1.20$, $F = 28.75$, $p < .001$; Wilcoxon Test, $p < .001$).

The participants in the control group performed at the same level on both the *physical* and *mental* inference tasks (ANOVA, $df = 1.19$, $F = 1.27$ $p = .27$; Wilcoxon Test, $p = .39$). They also did equally well on the A stories as on the B stories (ANOVA, $df = 1.19$, $F = 0.24$, $p = .63$; Wilcoxon Test, $p = .71$).

Prompt questions

As can be seen in Table 2, the AS group needed somewhat more prompt questions than the control group with regard to *physical* inference on the A stories, but this difference did not reach statistical significance ($df = 1.39$, $F = 5.16$, $p = .08$). This was confirmed by the analysis of covariance ($df = 1.37$, $F = 0.29$, $p = .60$). On the *mental* inference task, however, the AS group needed significantly more prompt questions than did the control group ($df = 1.39$, $F = 77.39$, $p < .001$). The covariance analysis demonstrated the same effect ($df = 1.37$, $F = 53.48$, $p < .001$).

On the B stories the covariance analysis showed that the AS group needed more prompt questions than the control group on *physical* inference ($df = 1.37$, $F = 6.03$, $p < .02$) as well as on *mental* inference performance ($df = 1.37$, $F = 34.92$, $p < .001$).

Within the AS group the need for prompt questions was markedly more apparent on *mental* than on *physical* inference performance. This was demonstrated on the A story set (ANOVA, $df = 1.20$, $F = 126.39$, $p < .001$) as well as on the B story set ($df = 1.20$, $F = 116.76$, $p < .001$).

The control group also needed significantly more prompts on *mental* than on *ph*ysical inference questions. This was the case on the A stories ($df = 1.19$, $F = 11.26$, $p = .003$) as well as on the B stories ($df = 1.19$, $F = 21.04$, $p < .001$).

Reaction time

Concerning the reaction times, the AS group used significantly more time than the control group on the *physical* inference task, both before the prompt questions had been asked ($df = 1.37$, $F = 15.79$, $p < .001$), and subsequently ($df = 1.37$, $F = 13.80$, $p < .001$). As expected, the difference between the groups was even more pronounced on the *mental* inference task. In this case the covariance analysis yielded a significant group difference before prompts ($df = 1.37$, $F = 21.84$, $p < .001$) as well as after the prompts had been given ($df = 1.37$, $F = 13.08$, $p < .001$). This lends some support to Bowler's finding (1997) about reaction time on standard second-order theory-of-mind tests.

Interestingly, *within* the AS group the participants used significantly longer reaction time on the *mental* inference tasks than on those involving *physical* inference (ANOVA, $df = 1.20$, $F = 30.35$, $p < .001$). This was also the case for the control group ($df = 1.19$, $F = 24.46$, $p < .001$). A comparison of the difference between the groups regarding reaction times on mental inference *minus* physical inference was also significant, both before prompt questions ($p < .001$) and after prompts had been given ($p = .0001$).

In the AS group the range in reaction time on the mental inference tasks was from a low of 66 to a high of 693 seconds, as opposed to 22–145 seconds among the controls. Table 2 lists the differences between the groups on mean total reaction time (inclusive/exclusive of the prompt questions) for physical and mental inference performances.

The long reaction times of the participants in the AS group can be seen as a consequence of their difficulties in answering the test questions, especially those involving mental states. The participants with AS seemed to need at least one, and quite often several prompt questions before being able to provide an answer. In general, for both groups the reaction times were longer for mental than for physical inference performance. As Table 2 shows, the difference was most evident for the clinical group as far as mental states were concerned.

Correlations

Among the participants in the AS group there was a significant relation between *age* and performance on *physical* inference on the A story set ($r = .46$, $p < .05$), but no correlation on the B story set ($r = .36$). The same variables were not significantly correlated among the controls ($r = .36$).

In the AS group there was a significant correlation between *age* and *mental* inference; ($r = .59$, $p < .01$) and ($r = .49$, $p < .05$) for the A and B set respectively. On the other hand there was no association between *VIQ* and performance on *physical* inference in either group (AS: $r = .03$; controls: $r = -.01$ and AS: $r = .10$; controls: $r = -.16$) for the A and B set, respectively. In the AS group there was no significant correlation between *mental* inference performance and *age* ($r = .38$).

Among the controls there was no significant relation between *VIQ* and *mental* inference on either story set.

Story type

As shown in Table 3, performance on the different story types varied in the AS group, primarily on the mental inference tasks. In some cases the physical inference tasks also presented problems for the AS participants. The differences between the two groups were much less evident on physical than on mental inference tasks, as shown above.

With respect to the different story types, certain stories proved more difficult for the AS group than did others. The story type *Figure of speech* appeared to be difficult for many of the AS participants. They tended to interpret the terms *the castle in the air* and *being in the same boat* literally. A *castle in the air*, many of them said, was an extremely high building. They were also puzzled by the expression *being in the same boat*, and explained this metaphor in many different and often idiosyncratic ways. In the control group only one participant answered this story type incorrectly.

Contrary emotions, *intentions* and *irony* were other story types that appeared to be difficult to understand for the participants in the AS group. In one of the *intention* stories about two girls on their way home from school, one of them asks her friend if she can carry her bag for a while. Her school friend answers yes, she can do that, but without understanding the intention behind the request and without signalling that she is ready to do so. Apparently, it was very difficult for the participants in the AS group to understand the girl's over-literal interpretation. Most of the AS persons attributed other motives to the girl for not helping her friend, in that they answered that she was provoking, cross, offended, busy, lazy, daydreaming or spiteful. Only four out of 21 participants in the AS group realised that the girl did not understand the intention behind the question. Interestingly, some of the controls also had problems with this story type.

On *contrary emotions* only 20% of the answers were totally correct in the AS group, whereas about 40% of the responses were partly correct. In one of the stories a woman is in an emotional dilemma between accepting a new, exciting job or staying at home caring for her 3-month-old son. In response to the question of why she says to her husband that she is happy about the new job, but says to her mother that she is not so happy, many of the participants in the AS group answered context-inappropriately that she was cross, lying to her

Table 3 Percentage of correct, partly correct and incorrect answers – after story type

	Physical inference			*Mental inference*		
Story type	*Answers:* % correct	% part. corr.	% incorrect	% correct	*Answers:* % part. corr.	% incorrect
AS group						
Lie	62	19	19	88.1	11.9	0
White lie	52.4	26.2	21.4	69.1	11.9	19.0
Figure of speech	71.4	28.6	0	38.1	7.1	54.2
Misunderstanding	71.4	23.8	4.8	50.0	31.0	19.0
Double bluff	95.2	4.8	0	57.2	23.8	19.0
Irony	81.0	11.9	7.1	57.2	4.8	38.1
Persuasion	81.0	7.1	11.9	76.2	16.7	7.1
Contrary emotion	95.2	2.4	2.4	28.5	31.0	40.5
Forgetting	73.8	21.4	4.8	95.2	2.4	2.4
Jealousy	95.2	4.8	0	78.6	19.0	2.4
Empathy	85.7	11.9	2.4	45.2	35.7	19.0
Intentions	92.6	4.8	2.4	50.0	7.1	42.7
Social blunders	100	0	0	50.0	37.7	14.3
Control group						
Lie	70.0	7.5	22.5	95.0	5.0	0
White lie	75.0	25.0	0	95.0	2.5	2.5
Figure of speech	95.0	5.0	0	92.5	5.0	2.5
Misunderstanding	80.0	20.0	0	100	0	0
Double bluff	100	0	0	97.5	0	2.5
Irony	90.0	2.5	7.5	82.5	5.0	12.5
Persuasion	90.0	5.0	5.0	95.0	2.5	2.5
Contrary emotion	100	0	0	87.5	7.5	5.0
Forgetting	72.5	22.5	5.0	100	0	
Jealousy	100	0	0	100	0	
Empathy	100	0	0	95.0	5.0	0
Intentions	100	0	0	67.5	7.5	25.0
Social blunders	100	0	0	92.5	7.5	10

husband or mother, respectively, or that she was calculating, had changed her mind, tried to flatter her mother, would avoid any discussion with her husband – or just saying 'life is like that sometimes'. Contrary to the participants in the control group, they did not recognise that she was in an emotional dilemma.

As expected, the persons in the AS group also had problems with the *irony* stories. Almost 40% of the answers were incorrect, and *irony* seems to be difficult for individuals with autistic spectrum disorder (Happé, 1991, 1994).

The AS participants were also troubled by the story type *empathy* in that 45% of the answers were fully correct and about 35% only partly correct. For *social blunders* half of the mental inference answers were fully correct and nearly 40% partly correct. On the stories comprising *misunderstanding* and *double bluff* the AS group performed relatively well, with about 80% fully/partly correct answers.

The participants in the AS group gave also many idiosyncratic responses to the story questions. For example, on one of the contrary emotions stories where a young mother was in an emotional dilemma about choosing between a good job that was offered her or staying at home caring for her 3-month-old baby, one of the participants of the AS group gave a wrong answer to the mental state justification question, answering that it was difficult to know whether she was *lying* either to her mother or to her husband when expressing her emotional dilemma. In one of the metaphor stories where two boys had wrecked their boat on a small island in the sea, one of them said that now we are in the same boat. On the question of what he meant by saying that, one of the AS participants said: 'How can I answer that bloody question? I hate explaining sayings.' In the story at the clockmaker's, a customer looking at an elegant gold watch suggests that the watch might be quite expensive; the owner confirms that by saying 'yes' and then immediately leaves her, failing to understand the intention behind the customer's question, namely to be informed about the exact price of the watch. One of the participants with a verbal IQ of over 120 answered the critical question – what she wanted to know – by saying that the clockmaker had answered her question! He then added that the customer was responsible for the breakdown of their communication, because she could have asked him more precisely. 'When one says it as she does, it is clear that one does not get much out of it', he added. He did not seem to understand that the clockmaker did not realise that there was an intention in her question.

Discussion

The aim of this study was to investigate whether children and adolescents with AS would be able to infer mental state concepts when the theory-of-mind elements were embedded in a 'naturalistic' context, and whether these persons would have more success with physical than with mental state inferences in the context of story comprehension.

As mentioned above, as many as 20 out of 21 AS participants passed the first- and second-order standard theory-of-mind tasks. Some researchers in the field have thought of second-order tests as 'complex' or high-level tests of theory of mind, but they are simply probes for skills at the 6-year-old level in this domain of functioning (Perner & Wimmer, 1985). Clearly, they are not suitable tests of whether adolescents and adults with AS, or any other clinical condition, have a fully functional theory of mind. The ability of people with autism and AS to put their theory-of-mind abilities into practice in real-life situations appears to be quite at odds with their test performance on standard theory-of-mind tests (Frith, Happé, & Siddons, 1994; Bowler, 1992). The present story test, we argue, may be closer to a real-life situation, in that the participants had to extract and integrate relevant information and process a range of mental states. The performance of the participants in the AS group on these 'advanced' tasks may also partly be due to failure to use context to guide their answers (Happé, 1994). So, for example, a participant with AS who explains a double bluff as a joke may be failing to use story context to inform his answer.

Could memory problems of the AS individuals be the underlying cause of the groups' differences? Some recent studies of high-functioning individuals with autism or AS indicate that the memory functions of these individuals may be relatively intact. Bowler, Gardiner, and Grice (2000) found no impairment of remembering in adults with AS, nor did the overall recognition differ as compared with normal control persons. Mottron, Morasse, and Belleville (2001) found no amnesic and semantic deficit in the memory of high-functioning individuals with autism relative to normal controls. In the present study, the differences between the groups on the mental inference tasks were robust, yielding a substantial *effect size*. In the test situation, relatively few of the participants in the AS group seemed to have difficulties with answering the control and physical state questions, in contrast to the mental state questions. Re-reading parts of the stories was mostly restricted to the last paragraph where the mental state inference components were embedded. It therefore seems unlikely that memory problems can account for the solid differences in mental state attribution found between the groups.

The major finding of the present study was that the participants with AS had significantly less success in inferring *mental* as compared with *physical states* on the *Stories from Everyday Life*. However, the possibility remains that the two sorts of test stimuli may not have been quite equivalent in difficulty; i.e., the *physical* inference questions may sometimes have been easier to answer than the *mental* ones. At least some of the participants in the AS group may have scored at ceiling level on the physical inference tasks. Nonetheless, the differences between the groups were significant on the two types of task, and we argue that they can hardly be explained by a ceiling effect alone. Moreover, the within-group difference among the AS participants on the *physical* and *mental* inference tasks, respectively, was also solid.

On Happé's *Strange stories*, the participants with AS in the present study showed considerable difficulties with the mentalistic questions as compared with the control group, and the problems they showed were largely of the same kind as those manifested on the *Stories from Everyday Life*. The findings using Happé's *Strange stories* test lend support to Happé study (1994), also replicated by Jolliffe and Baron-Cohen (1999), finding that high-functioning individuals with autism or AS are impaired on this task. Interestingly, relatively high relations were found between performance on the mental inference tasks of the *Stories from Everyday Life* and Happé's *Strange stories*, suggesting that the theory-of-mind elements embedded in a story context may tap similar mental abilities.

The relatively high association between intelligence level and mental inference performances reported in the present study might lead some researchers in the field to conclude that the mentalising deficit is strongly related to intellectual competence, and that it does not constitute a distinct disorder (Prior et al., 1998).

One may argue against this view. The fact that the clinical group used significantly longer reaction times, and needed significantly more prompt questions than the controls in solving the mental inference tasks, may also highlight their problems with interpreting social communication. Also, many participants in the clinical group revealed idiosyncratic ways of interpreting everyday situations, especially situations where mental states were involved. Amongst the most striking *qualitative* finding was the AS persons' tendency to interpret events literally, when a mental state interpretation would be the most common option. Bauminger and Kasari (1999) also found that their high-functioning persons with autism, who passed a second-order standard theory-of-mind test, responded in significantly more incorrect, irrelevant and idiosyncratic ways on mental state inferences, as compared with normally developing controls.

As mentioned above, few previous studies have collected data on speed of problem solving on theory-of-mind tasks. The current recording of reaction

times showed that many of the AS participants were slow and longwinded in the test situation (Wing, 1998), and significantly slower than the controls, especially on mental inference performance. Hermelin and O'Connor (1985), who invoked the concept of 'logico-affective' states, have argued that some individuals with autistic spectrum disorders are able to use cognitive mechanisms to deal with problems that are usually dealt with by affective processes. Since such problems are usually encountered in the realm of social interactions, the use of cognitive rather than affective processes could, as Bowler (1992) suggests, result in a slowing of responses, and a disruption of the subtle timing of social interactions, thereby making the person seem odd to other people.

This view of autistic social impairment fits well with clinical accounts, and echoes Bruner and Feldman's (1993) description of interactions of individuals with AS. These interactions are likened to conversing with someone who is calculating his next move, or with someone who is working out a complicated maths problem. However, to what extent the longer reaction times among AS persons are related to their mentalising problems, or at least partly represent a separate phenomenon, is as yet not clear.

In the current study, some of the participants in the AS group provided correct responses to the mental inference tasks of the *Stories from Everyday Life*. At first sight this may indicate that these individuals have an intact mentalising mechanism. It is possible that those relatively able individuals with AS, who answered a substantial proportion of the test questions correctly, may have a less severe impairment of their theory of mind. Actually, very few, if any, answered the test questions as did the normal controls. Those who provided more or less correct responses often needed prompt questions because they either did not understand the mental state question or provided inappropriate responses. Some of them gave also additional, but context-inappropriate, answers to their already correct answers. Certainly, the understanding of mental states developed by even the most able participants in the AS group may differ markedly from the spontaneous and effortless mentalising abilities of the participants in the control group. As yet, no single case has been reported of a child with autism or AS who attributed mental states in standard tests at the early age when young children normally show such an ability, and they also require considerably higher verbal abilities to solve such tasks (Frith & Happé, 1999).

The current study indicates that individuals with AS, at least the most able of them, are not quite lacking in ability in attributing mental states, but they do it differently as compared to normally developing individuals. Many of them seemed to process social events in rather idiosyncratic ways, in contrast to the spontaneous and intuitive interpretation of social events made by the controls. This might be related to their literal thinking and to their slow and 'digital' way of processing social events, indicating a somewhat different 'cognitive style' (Happé, 1999) or 'difference' in mentalising abilities (Baron-Cohen, 2000). There is some experimental evidence of a relation between young children's theory of mind and social understanding in a naturalistic setting (Astington & Jenkins, 1995) and the same type of relation also seems to exist in individuals with autism or PDD (Frith et al. 1994; Hughes, Soares-Boucaud, Hochmann, & Frith, 1997). Such a relation may have – treating autism and AS as two faces of the same disorder – significant consequences for the AS persons' functioning in everyday social life, since social understanding in naturalistic settings requires mental flexibility and subtle timing of social responses.

Correspondence to

Nils Kaland, Høgskolen i Lillehammer, Serviceboks, 2626 Lillehammer, Norway; Email: nils.kaland@hil.no

References

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

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: Author.

Asperger, H. (1944) Die 'Autistischen Psychopathen' im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, *117*, 76–136.

Astington, J.W., & Jenkins, J.M. (1995). Theory of mind development and social understanding. *Cognition and Emotion*, *9*, 151–165.

Attwood, T. (1998). *Asperger's syndrome. A guide for parents and professionals*. London: Jessica Kingsley Publishers.

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

Baron-Cohen, S. (2000). Is Asperger syndrome/high-functioning autism necessarily a disability? *Development and Psychopathology*, *12*, 489–500.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Roberts, 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*, 813–822.

Baron-Cohen, S., Wheelright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the mind in the Eyes' Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

Bauminger, N., & Kasari, C. (1999). Brief report: Theory of mind in high-functioning children with autism.

Journal of Autism and Developmental Disorders, *29*, 81–86.

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

Bowler, D.M. (1997). Reaction times to mental state and non-mental state questions in false belief tasks by high-functioning individuals with autism. *European Child and Adolescent Psychiatry*, *6*, 160–165.

Bowler, D.M., Gardiner, J.M., & Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *30*, 295–304.

Bruner, J., & Feldman, C. (1993). Theories of mind and the problem of autism. In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.), *Understanding other minds – perspectives from autism*. Oxford: Oxford University Press.

Carruthers, P. (1996). Autism as mind-blindness: An elaboration and partial defence. In P. Carruthers & P.K. Smith (Eds.), *Theories of theories of mind*. Cambridge: Cambridge University Press.

Dahlgren, S.O., & Trillingsgaard, A. (1996). Theory of mind in non-retarded children with autism and Asperger's syndrome. A research note. *Journal of Child Psychology and Psychiatry*, *37*, 759–763.

Ehlers, S., & Gillberg, C. (1993). The epidemiology of Asperger Syndrome. A total population study. *Journal of Child Psychology and Psychiatry*, *34*, 1327–1350.

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it to be autistic? *Mind and Language*, *14*, s. 1–22.

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

Ghaziuddin, M., & Butler, E. (1998). Clumsiness in autism and Asperger syndrome: A further report. *Journal of Intellectual Disability Research*, *42*, 43–48.

Gillberg, C. (1997). Barn, ungdommar och vuxna med Asperger syndrom – normala, genier, nørder? Stockholm: Bokförlaget Cura AB.

Happé, F.G.E. (1991). The autobiographical writing of three Asperger syndrome adults: Problems of interpretation and implications for theory. In U. Frith (Ed.), *Autism and Asperger syndrome*. Cambridge: Cambridge University Press.

Happé, F.G.E. (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*, 129–154.

Happé, F.G.E. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, *3*, 216–222.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The awkward moment test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, *30*, 225–236.

Hermelin B., & O'Connor, N. (1985). Logi-co-affective states and non-verbal language. In E. Schopler & G.B. Mesibov (Eds.), *Communication problems in autism*. New York: Plenum Press.

Hughes, C., Soares-Boucaud, I., Hochmann, J., & Frith, U. (1997). Social behaviour in pervasive developmental disorders: Effect of informant, group and 'theory of mind'. *European Child and Adolescent Psychiatry*, *6*, 191–198.

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

Kaland, N. (2000). *Evne til mentalisering ved Asperger syndrom. En studie av mentalisering, 'sentral koherens' og eksekutive funksjoner*. Unpublished doctoral dissertation, University of Oslo, Institute of Psychology.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kleinman, J., Marciano, P.L., & Ault, R.L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*, 29–36.

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*, 831–846.

Klin, A., Volkmar, F.R., Sparrow, S.S., Ciccietti, D.V., & Rourke, B.P. (1995). Validity and neuropsychological characterization of Asperger syndrome – convergance with nonverbal learning-disability syndrome. *Journal of Child Psychology and Psychiatry*, *36*, 1127–1140.

Mottron, L., Morasse, K., & Belleville, S. (2001). A study of memory functioning in individuals with autism. *Journal of Child Psychology and Psychiatry 42*, 253–260.

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

Ozonoff, S., Rogers, S.J., & Pennington, B.F. (1991). Executive function deficits in high-functioning autistic children: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, *32*, 1081–1106.

Perner, J., & Wimmer, H. (1985). 'John thinks that Mary thinks that….' Attribution of second-order belief by 5–10 years old children. *Journal of Experimental Child Psychology*, *39*, 437–471.

Perner, J., Frith, U., Leslie, A.M., & Leekam, S.R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, *60*, 689–700.

Prior, M., Leekam, S., Ong, B., Eisenmajer, R., Wing, L., Gould, J., & Dove, D. (1998). Are there subgroups within the autistic spectrum? A cluster analysis of a group of children with autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, *39*, 893–902.

Shrout, P. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, *7*, 301–317.

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders*, *24*, 577–586.

Volkmar F.R., Klin, A., Schultz, R.T., Rubin, E., & Bronen, R. (2000). Asperger's syndrome. *American Journal of Psychiatry*, *157*, 262–267.

Wechsler, D. (1998). Wechsler Intelligence Scale for Children (WISC-III). Danish translation and adaptation. Danish Psychological Publishers.

Wing, L. (1981). Asperger's syndrome. A clinical account. *Psychological Medicine*, *11*, 115–129.
Wing, L. (1991). The relationship between Asperger's syndrome and Kanner's autism. In U. Frith (Ed.), *Autism and Asperger syndrome*. Cambridge: Cambridge University Press.
Wing, L. (1998). The history of Asperger syndrome. In E. Schopler, G.B. Mesibov, & L.J. Kunce (Eds.), *Asperger syndrome or high-functioning autism*. New York: Plenum Press.
World Health Organization (1993). The ICD-10 classification of mental and behavioural disorders. Diagnostic criteria for research. Geneva: Author.

Manuscript accepted 15 November 2001

Appendix 1
Two examples of *Stories from Everyday Life*

Castles in the air

The architect Ken Peterson is known as a person rich in ideas. He works with Solnes, a master builder who has his office in the town. He goes to Solnes almost daily with new ideas about how to build bigger and better buildings.

The idea-rich architect uses steel and glass as construction materials, because they are the materials that can give the most protection against storms and bad weather. With these materials it is possible to build fine, big buildings. Wooden material and roof tiles are well suited for the construction of normal single-floored dwellings, he says.

Many of the people who hear of Peterson's many building plans regard them as quite unrealistic. Solnes, the master builder, is also normally sceptical of the architect's ideas. One day Peterson arrives and says that he has begun drawing plans for the town's new city hall. He will build it high, he says, 35 floors – because this will save on land area. Solnes, the master builder, thinks that this and a number of Peterson's other recent ideas are totally unrealistic. Solnes says: 'Peterson, now I think you are building castles in the air.'

Questions

1. What is an architect?
2. What is architect Peterson known for?
3. Who does he work with?
4. What is a builder?
5. What do most people think of Peterson's ideas?
6. What does master builder Solnes think of them?
7. Why doesn't architect Peterson use wooden material and roof tiles when projecting high buildings? (PI)

Answer:.....

8. How many floors does Peterson intend the new Town Hall to have?
9. What is Solnes' opinion of building so high?
10. Does Solnes really mean that Peterson is planning to build a castle made only of air?

YES/NO/DON'T KNOW

11. What does Solnes mean when he says that Peterson builds castles in the air? (MI)

Answer: .....
Prompt: .....

Fido

The Hanson family, Mrs Elsie, her husband Gerald and their children, Emma and Dan, have a large, kind, fowl-hunting dog called Fido. Both Emma and her brother Dan are very fond of Fido. Every day Fido sits on the doorstep, looking out for Emma and Dan when they come home from school, and wags his tail when he sees them.

When Emma and Dan's mother was young, she was bitten by a dog. Since then she has never liked dogs, and she is not particularly fond of Fido. Also, she complains that Fido regularly runs after birds in the muddy ground close by.

When the dog isn't outside, it is usually to be found in the kitchen. Elsie has to wash the kitchen floor almost daily. Even though she knows that her husband and her children are fond of the dog, she has several times said to her husband that she would like to get rid of Fido. Her husband is against this, especially because the children are extremely fond of Fido.

Emma has asthma, and suffers sometimes from asthma attacks, generally when she is at school. One day she has an attack and is almost unable to breathe. Luckily, she has her asthma spray in her school bag, so she soon recovers. When her mother hears about this she says to her husband: 'I am quite sure that Emma's asthma attack was caused by an allergy to dogs, and that this is Fido's fault. It is therefore time to get rid of this dog, before it ruins Emma's health!'

Questions

1. What is the Hanson family's dog called?
2. What do Emma and Dan think of Fido?
3. Why does Fido sit and wait for Emma and Dan to return from school?
4. What does Emma and Dan's mother think of dogs?
5. Why does she wash the kitchen floor almost daily? (PI)

Answer: .....
Time: .....

6. What does she want to do with Fido?
7. What do her husband and children think of this?
8. What kind of illness has Emma?
9. What does Emma's mother say to her husband after hearing about Emma's asthma attack at school?
10. Where is Emma when she normally has her attacks?

11. Is Fido normally present when Emma has her asthma attacks?
YES/NO/DON'T KNOW
12. Why does Emma's mother say that Fido is the cause of Emma's asthma attacks, even though she has her attacks when the dog is not present? (MI)
Answer: .....
Time: .....
Prompt: .....

Appendix 2

Inter-rater kappa values for physical and mental inference on *Stories from Everyday Life* (SEL)

	Stories from Everyday Life – A		*Stories from Everyday Life* – B	
Story type	Physical inference	Mental inference	Physical inference	Mental inference
1 Lie	.86	.70	.82	.76
2 White lie	.46	.85	.62	.91
3 Figure of speech	.73	.90	.93	.86
4 Misunderstanding	.42	.84	.65	1.0
5 Double bluff	1.0	.88	.79	.87
6 Irony	.82	1.0	.54	.90
7 Persuasion	.82	.86	.71	.60
8 Contrary emotions	.65	.92	1.0	.86
9 Forgetting	.65	.79	.62	1.0
10 Jealousy	.66	1.0	.48	1.0
11 Empathy	.79	.74	.82	.74
12 Intentions	1.0	.78	1.0	.96
13 Social blunders	1.0	.81	1.0	.78
Mean kappa values	.76 (.41–1.0)	.85 (.69–1.0)	.77 (.48–1.0)	.87 (.59–1.0)