J Autism Dev Disord (2008) 38:905–918
DOI 10.1007/s10803-007-0462-5

ORIGINAL PAPER

Mind-Reading in Young Adults with ASD: Does Structure Matter?

Koen Ponnet · Ann Buysse · Herbert Roeyers · Armand De Clercq

Published online: 10 October 2007

Abstract This study further elaborates on the mind-reading impairments of young adults with autism spectrum disorder (ASD). The hypothesis is that differences in mind-reading abilities between subjects with ASD and control subjects become more apparent when they have to infer thoughts and feelings of other persons in a less structured or more chaotic conversation, than when they have to do so in a more structured conversation. Conform to the empathic accuracy design, subjects viewed two videotaped interactions depicting two strangers and attempted to infer thoughts and feelings. One of the videotaped conversations was less structured than in the other. The results underscore the significance of structure to the mind-reading abilities of young adults with ASD.

Keywords Autism · ASD · Empathic accuracy · Mind-reading

Mind-Reading in Young Adults with ASD: Does Structure Matter?

People with autism or an autism spectrum disorder (ASD) experience profound social deficits, which tend to persist over time. Even the most able individuals with ASD have difficulties in understanding and responding appropriately to people's feelings or emotions (Howlin 1997; Lord 1993; Zager 1999). A psychological theory of autism that has been proposed to account for these social problems is the "theory of mind" hypothesis. This hypothesis suggests that the impairments of subjects with autism or ASD can be explained in terms of a failure to attribute mental states (such as beliefs, desires or intentions) and feelings to oneself and others, which results in socially and communicatively inadequate behaviour (Baron-Cohen et al. 2000). Without a theory of mind, people are unable to develop a normal understanding that other people have mental states and feelings. However, one of the difficulties in assessing the "theory of mind" hypothesis of autism is that there is no clear single statement of the hypothesis (Boucher 1989). In its narrow use, the term refers only to the ability to impute cognitive or volitional states to others (Premack and Woodruff 1978). In its broader use, the terms also refers to mind-reading, that covers more directly on-line processing on mental state information including both verbal and non-verbal cues, thoughts and feelings. In the present study, the focus is on the latter.

Over the past decades, an increasing amount of research was devoted to the development of mind-reading tasks for use with children with ASD (Happé 1994a; Sodian and Frith 1992). With a few exceptions (e.g. Bauminger and Kasari 1999; Dahlgren and Trillingsgaard 1996; Prior et al. 1990), these studies largely support the theory of mind hypothesis. However, the classic mind-reading tasks are not ideal for testing older, high-functioning individuals with ASD (Rutterford et al. 2002). The last few years, research began to focus on the development of more advanced adult mind-reading tests (Baron-Cohen et al. 1997, 2001; Rutterford et al. 2002). These advanced tasks

K. Ponnet (✉) · A. Buysse · H. Roeyers
Department of Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
e-mail: Koen_Ponnet@yahoo.com

Present Address:
K. Ponnet
Research Centre for Longitudinal and Life Course Studies, University of Antwerp, Antwerp, Belgium

A. De Clercq
Department of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium

have a more mature and challenging content, which distinguishes them from mind-reading tasks developed for younger subject populations (Heavey et al. 2000). However, the ecological validity of most of the advanced mind-reading tasks is still debatable, because the greater part of these tasks makes use of static stimuli (i.e. verbal *or* non-verbal stimuli). In order to measure the mind-reading of adults with ASD in contexts that approximate closely the real social world, several authors have called for more dynamic tasks, in which verbal *and* non-verbal stimuli (e.g. videomaterial) are used (Baron-Cohen et al. 2001; Kleinman et al. 2001; Roeyers et al. 2001). Until now, only few tasks were designed to examine the mind-reading or emotion recognition abilities in adults with ASD, through both visual and auditory stimuli, using videomaterial. However, the rare studies that made use of videotapes proved to be sensitive in detecting mind-reading difficulties in adults with ASD compared to typically developing controls. Notable examples are the Awkward Moment Test (Heavey et al. 2000), the Cambridge Mind-reading (CAM) face-voice battery (Golan et al. 2006) and the Movie for the Assessment of Social Cognition (MASC) (Dziobek et al. 2006).

A validated research paradigm that offers the opportunity for a systematic investigation of the mind-reading in a laboratory situation is the empathic accuracy design of Ickes and colleagues (Ickes 1993; Ickes et al. 1990a). Empathic accuracy is the extent to which people are successful in the "everyday mind-reading" they do whenever they attempt to infer other people's thoughts and feelings (Ickes 1993). The term "empathy" comes from mainstream social psychology and has as such been defined in different ways by different authors (see Baron-Cohen 2003; Eisenberg et al. 1997; Preston and de Waal 2002). Opinions differ greatly on the matter whether "empathy" should be viewed as unidimensional versus multidimensional, as cognitive versus emotional, or as a global skill versus a set of component subskills (Ickes et al. 1997, p. 283). It is however widely accepted that the term "empathy" includes both a cognitive and an emotional component (Davis 1994). While the cognitive component refers to the capacity to understand other's internal states, the emotional component refers to the emotional response resulting from the recognition of other person's emotional state or condition (Eisenberg et al. 1991). However, similar to mind-reading tasks, the focus of the empathic accuracy design is merely on one component of empathy, i.e. recognising the mental states of others. Over the years, the empathic accuracy paradigm has consistently proved to be a reliable and valid way to measure peoples' inference abilities (Buysse and Ickes 1999; Ickes et al. 1990b; Marangoni et al. 1995; Simpson et al. 1995; Stinson and Ickes 1992).

In two recent studies (Ponnet et al. 2004; Roeyers et al. 2001), we tested adults with ASD and control adults using a naturalistic Empathic Accuracy Task. In this task, individual participants each attempted to infer the thoughts and feelings of a target person, while viewing a videotape of the target person in a naturally occurring conversation with a stranger. In both studies, the Empathic Accuracy Task was able to distinguish between the adults with ASD and the control adults, although a significant between-group difference was found only for one of the empathic accuracy videotapes in favour of the control adults. The mixed results can be explained by the different nature of both videotape tasks. Although we had not the intention to manipulate the videotapes' degree of structure, the social interaction in the first videotape was more structured than that in the second one.

Since the empathic accuracy paradigm has proven to be a reliable way to measure adults' inference abilities as well as to distinguish adults with ASD from typically developing adults, we want to test a more fine-grained hypothesis. More specifically, the present study attempts to explore the possible significance of structure to the empathic accuracy of subjects with ASD. As is well known, subjects with ASD prefer activities and situations that are more structured and provide clear, explicitly stated rules (Bauminger and Shulman 2003; Howlin 1997; Zager 1999). Furthermore, many authors agree that most subjects with ASD benefit so much from structure that their most severe symptoms will not be shown in an environment with a high degree of structure (Howlin 1997; Gillberg et al. 1996; Mesibov 1992). However, structure has seldom explicitly been manipulated in empathic situations. This study is related to this issue.

In the present study, we developed a new Empathic Accuracy Task in which young adults with ASD have to infer the thoughts and feelings of other persons. The procedure used was the standard stimulus empathic accuracy paradigm (Marangoni et al. 1995), in which individual participants (i.e. subjects with ASD and control subjects) each view two videotaped "getting acquainted" conversations between two strangers and attempt to infer the thoughts and feelings of the same set of target persons. We manipulated the structure of both videotapes in such way that one videotape was more structured than the other, and we suggest that the inference impairments of young adults with ASD may only become apparent when confronted with the less structured conversation.

In addition to the structure manipulation, the present Empathic Accuracy Task extends our previous studies in two ways. First, we studied content accuracy as well as valence accuracy of subjects with ASD and control subjects, whereas in our previous studies we only calculated content accuracy. Content accuracy refers to the degree to which the perceiver's description of the target's thoughts or feelings matches the content of the specific thought or feeling reported by the target. Valence accuracy refers to the degree to which the perceiver's inferences about the emotional tone of the target's actual thoughts or feelings matches the actual valence of the specific thought or feeling reported by the target (Ickes et al. 1990b). Second, we can expect that variables other than structure might influence the mind-reading abilities of subjects with ASD as well. In a conversation, the subjects' verbal and non-verbal behaviours provide information about their inner thoughts and feelings. So, subjects who are good interpreters of behavioural cues will be better at inferring other people's thoughts and feelings. Although many subjects with ASD are aware that verbal and non-verbal cues play a crucial role in social interactions, it might be difficult to them to interpret these cues correctly (Howlin 1997; Klin et al. 2002; Klin and Jones 2006). Therefore, we studied both the overt behaviour of the videotaped strangers as well as the covert thoughts and feelings of the interacting targets.

In the present study, we expect significant differences between adults with ASD and control adults when they have to infer the content and valence of other persons' thoughts and feelings in a more chaotic and less structured conversation. However, no significant between-group differences are expected when subjects have to infer the content and valence of other persons' thoughts and feelings in a more structured conversation. In addition, we expect to get a clearer picture of the empathic accuracy scores of subjects with ASD when more refined content accuracy scores are calculated. More specifically, we expect the content accuracy scores to decrease when subjects have to infer thoughts and feelings of other persons that are more difficult, abstract and incongruent with the videotaped persons' verbalisation and their direction of gaze, with subjects with ASD having more difficulties than control subjects. Based on a study of Golan et al. (2006) that suggests that subjects with ASD do not find positive emotions easier to recognise than negative ones, we expect that the inference ability of subjects with ASD will not differ along the affective tone (i.e. positive, negative or neutral) of the videotaped targets' thoughts and feelings.

Methods

Participants

Two groups of normally intelligent adolescents and adults participated in the present study: 22 male subjects with ASD and a control group of 22 typically developing male subjects. The subjects were recruited with the help of parent associations and schools. All subjects with ASD had been diagnosed by a multidisciplinary team of experienced clinicians and fulfilled established DSM-III-r or DSM-IV criteria for autism, Asperger syndrome or PDD-NOS (APA 1987, 1994), depending upon which was most recent at the time of diagnosis. The mean general level of intelligence was within the normal range. Only subjects with a Verbal IQ of at least 75 and a minimum chronological age of 15 years were included in the sample. The typically developing control subjects were matched meticulously on a one-to-one basis on sex, chronological age and IQ. All subjects were administered 11 of the 14 subtests of the WAIS-III (Wechsler 2000), so that the Total IQ, Verbal IQ and Performance IQ could be estimated. Further background information for both groups is presented in Table 1.

Table 1 Descriptive characteristics of the sample

	ASD		Control		$F(1,43)$
	M	SD	M	SD	
Chronological age	19.32	3.48	19.26	3.15	0.00
Verbal IQ	98.05	13.31	96.32	11.32	0.21
Performance IQ	89.45	16.81	95.45	13.66	0.20
Total IQ	93.82	13.66	95.55	12.69	0.67

Materials

Empathic Accuracy Stimulus Tapes

The stimulus materials consisted of two videotapes of two volunteers, with varying level of structure. For each of the videotapes, two opposite-sex individuals (targets), previously unacquainted with each other, were scheduled to come to our research centre at the same time. They were asked to participate in a study on "people's perceptions of others during different situations". They were told in advance that they were about to meet another stranger and that their interaction would be videotaped for use in future research. Once they arrived at the research centre, the two targets were brought to different waiting areas so that they would not meet and interact before the session began. They were then brought together.

In the *first* videotape, the research assistant told the participants they had to fill in some inquiry forms. However, when the research assistant reached for the inquiry forms, he "discovered" that some of the copies were not well printed. In order to get some proper copies, the experimenter left the room. Promising to return in a few minutes, he suggested both participants to seek acquaintance with each other during his absence. At this point, a concealed video camera was activated and recorded the initial conversation between the two targets.

In the *second* videotape, the research assistant explained the targets they were to play a board game together, but that the rules of the game require that both participants get to know each other personally before starting the game. In order to become acquainted with each other in a decent and less stressful manner, the experimenter proposed to leave the room for approximately 8 min. Before leaving, he gave the targets an eight-point list with questions they surely have to know from each other (see Appendix A), and encouraged the targets to get acquainted with each other as closely as possible. At the moment the experimenter left the room, a second research assistant activated a concealed video camera.

In both videotapes, the research assistant returned after approximately 8 min and partly debriefed the targets. The research assistant further informed the targets that their written consent was required for the tape to be used. The stimulus tapes consisted of two initial conversations between two strangers. The targets were all white, 18 or 19 years old (within the age-range of the participants), and college students. We presented the preparation of the stimulus tapes in the above-mentioned contexts for several reasons. First, both contexts served as a plausible reason to invite the targets to come to the research centre and enabled us to videotape them unobtrusively. Second, the explicit instruction in video 2 to get to know each other more personally increased the possibility of a more structured conversation between the targets, whereas the absence of such an explicit instruction in video 1 did not. Third, the questions of the eight-point list (see Appendix A) can be considered as highly typical questions for an initial conversation of the getting acquainted type. Therefore, instructing the targets in video 2 to base their conversation on the list was a guarantee that the conversation of the targets would be one of the "getting acquainted" type and also structured the conversation. The conversation of videotape 2 followed a more stereotyped and conventional pattern than that of videotape 1, and the dyads of videotape 2 openly referenced the prompted questions.[1] Finally, as will be described in the procedure, the list was presented to the participants as a guideline, which also enhances the structuralisation.

Immediately following the videotape session, the targets were asked to view their videotape and to make a written record of all the unexpressed thoughts and feelings they remembered having had during the acquaintance. The targets were each seated at a separate small table in one of two respective areas of the room, where each of them had a clear view of a single, large TV monitor. Following the procedure of previous studies (Buysse and Ickes 1999; Ickes et al. 1990a; Roeyers et al. 2001), the videotape was activated and the targets were instructed to say "stop" at each point during the interaction when they remembered having had a specific thought or feeling. At each of those stops, the targets were asked to write down on a standardised thought/feeling coding form (a) the time when the thought or feeling occurred (as displayed by a digital clock on the under-right corner of the videotape), (b) whether the entry was a thought or a feeling, (c) the specific content of the thought/feeling entry and (d) whether the entry was positive, negative, or neutral in its overall affective tone. The targets were instructed to provide an accurate, honest and complete account of all the thoughts and feelings they remembered having had. The importance of being completely candid was emphasised. In addition, the targets were guaranteed that their actual thought/feeling entries would never be shown to their interaction partner and that their entries would be treated anonymously. On the other hand, the instruction cautioned them to report only those thoughts and feelings they distinctly remembered having had during the interaction period and not to report any new thoughts and feelings that occurred to them while viewing the videotape. Although it might seem naïve to assume that people are capable of accurately and honestly reporting their past thoughts and feelings, considerable evidence for the construct validity of this method of thought-feeling assessment is available to support this assumption (see Ickes et al. 1986).

Video 1 contained 43 thought/feeling entries (22 females/21 males) and video 2 contained 74 thought/feeling entries (36 females/38 males). After recording the thoughts and feelings, the targets were debriefed more thoroughly. All targets agreed to sign a release form allowing their videotape and thought/feeling data to be used as stimuli in the present study.

Because of the large number of thoughts and feelings in both videotapes, we reduced the number of thoughts and feelings in a random stratified way to 28 thought/feeling entries per videotape, so that each videotape contained 14 males and 14 females thought/feeling entries. Video 1 lasted for 7 min 31 s and video 2 lasted for 6 min 53 s.

A chi-square revealed that the numbers of thought/feeling entries indicated by the videotaped targets as positive (10 in video 1 and 11 in video 2), neutral (nine in video 1 and eight in video 2) or negative (nine in video 1 and nine in video 2) did not differ significantly between both videotapes, $\chi^2(1) = 0.11$.

Sample thought/feeling entries reported by the male target in the first videotape were: "I have to say

[1] Sample sentences uttered by the targets in the second videotape were: "We have to know each other very well, so tell me everything about your studies", "What's the next question on the list?" or "Do we share some interests?" (see Appendix A).

something", "What a lovely girl", "Shall I say something funny?". Sample entries reported by the female target in the first videotape were: "What is he saying?", "I have to ask him something?". Sample thought/feeling entries reported by the male target in the second tape were: "Ouch, we have to get to know each other personally", "Which are our mutual hobbies?", "Wow, I promote myself a lot". Sample entries reported by the female target in the second videotape were: "Shall I laugh to decrease the tension?", "I feel more at ease than only a minute ago".

As mentioned above, video 2 was more structured than video 1. To check the videotapes' level of structure, 60 typically developing persons (30 males/30 females) between 16 and 32 years old (M = 22.60, SD = 3.39) were invited to come to our laboratory. All of them were students who were naïve to the aims of the study. They were instructed to view the videotapes with the order of videotape presentation counterbalanced between subjects. After each videotape, the subjects had to rate on a ten-point scale, ranging from 1 (not at all typical) to 10 (very typical), to what degree 40 items were typical for the videotaped interaction. Among these 40 items, 13 items[2] assessed how structured each videotape was. The internal consistency (Cronbach's alpha) of the 13 items provided by the 60 subjects was 0.80 for video 1 and 0.90 for video 2. Then, mean scores were calculated for each participant and for each videotape, by adding up for each videotape the participant's scores on the 13 items and by dividing this aggregate score by the total number of items. The mean structure of video 1 was 4.94 (SD = 1.14) on a total score of 10, and the mean structure of video 2 was 7.04 (SD = 1.32). A paired t-test revealed a significant difference between both videotapes, $t(59) = -12.06$, $p < 0.001$, indicating that video 2 was more structured than video 1.

Procedure

All participants were tested individually in a quiet laboratory room at the University. Each participant was tested in two phases. In the first phase, the WAIS-III was administered. Approximately 2 weeks later, subjects were invited to come for a second time, during which they were instructed to view the empathic accuracy stimulus tapes with the order of videotape presentation counterbalanced between the participants. While viewing the videotape, the tape was paused at precisely those moments during which a target had recorded a specific thought or feeling. The stimulus tapes presented to the subjects were paused automatically with the use of VIDANN (De Clercq et al. 2001). Whenever, the videotape was paused, the subjects were asked to make inferences about the specific content of the unexpressed thought/feeling entries and to write down on a standardised coding form (a) whether the entry was presumed to be a thought or feeling, (b) the specific content of the thought/feeling entry and (c) whether the entry was presumed to be positive, negative, or neutral in its overall affective tone. Before each videotape, the participants were presented a powerpoint-presentation with minimal background information about the construction of the videotapes and, with regard to video 2, the eight-point list (see Appendix A). Providing the participants with the eight-point list, guaranteed that the structure of the conversation was clear to them. Furthermore, to ensure that the subjects clearly understood the procedure of the Empathic Accuracy Task, an advanced preparatory session was given.

[2] The 13 items were synonyms or were connected semantically with each other (e.g. "structured", "organized", "surveyable", "coherent", and "predictable").

Measures

Empathic Accuracy Task: Actual Thought/Feeling Measures

As suggested by Ickes et al. (1990a), the subjects' thought/feeling entries were coded by five independent coders. The coders watched the videotapes that were stopped each time one of the targets had reported having had a specific thought or feeling. The coders were instructed to rate for each thought/feeling entry how difficult they thought it was to infer the specific content of each thought/feeling entry on a seven point scale ranging from 1 (very easy) to 7 (very difficult). Because the internal consistency (Cronbach's alpha) of the five judges was 0.83 for all thought/feeling entries, the coders' scores were averaged for each thought/feeling entry. A t-test for dependent measures revealed no significant difference between the mean difficulty of the thought/feeling entries of video 1 (M = 4.17, SD = 1.15) and the mean difficulty of the thought/feeling entries of video 2 (M = 4.05, SD = 1.20), $t(139) = 0.51$. Then, based on the quartile distribution of both videotapes' thought/feeling entries, the thought/feeling entries were divided into (a) the 25.0% most easy-to-infer thought/feeling entries, (b) the 48.2% moderate difficult-to-infer thought/feeling entries and (c) the 26.8% most-difficult-to-infer thought/feeling entries.

Similarly, the same five coders rated for each thought/feeling entry how concrete or abstract the thought or feeling was on a seven-point scale ranging from 1 (very concrete) to 7 (very abstract). Because the interrater reliability (Cronbach's alpha) was 0.77 for all thoughts and feelings, the coders' scores were averaged for each thought/feeling entry. The mean abstractness of the thought/feeling entries of video 1 was 5.16 (SD = 1.07) and the mean abstractness of the thought/feeling entries of video 2 was 5.09 (SD = 0.65). A *t*-test for dependent measures revealed no significant difference between both videotapes, $t(139) = 0.60$. Following the above-mentioned procedure, the thought/feeling entries were divided into (a) the 25.0% most abstract thought/feeling entries, (b) the 51.8% moderate abstract thought/feeling entries and (c) the 23.2% most concrete thought/feeling entries.

Finally, the same five coders assessed for each target's thought/feeling entry whether or not the content of each thought/feeling entry was congruent with (a) the verbalisation of the target and (b) with the target's direction of gaze on a four-point scale ranging from 1 (not at all congruent) to 4 (very congruent). The interrater reliability (Cronbach's alpha) was 0.80 for all thoughts and feelings that were congruent with the verbalisation of the target, with "Cronbach's alpha if item is deleted" varying between 0.71 and 0.80. We recoded the assessment of the coder with the lowest "Cronbach's alphas if item is deleted" into a binary scale (not congruent/congruent) and used it in the further of the study. The interrater reliability (Cronbach's alpha) was 0.82 for all thoughts and feelings that were congruent with the target's direction of gaze, with Cronbach's alphas if item is deleted varying between 0.77 and 0.81. The assessment of the same coder as above had the lowest "Cronbach's alphas if item is deleted". We recoded this assessment into a binary scale, which is used in the further of the study. In the first videotape 8 thought/feeling entries (28.6%) of the targets were congruent with their verbalisation. In the second videotape, 15 targets' thought/feeling entries (53.6%) were congruent with their verbalisation. A chi-square revealed that the difference between both videotapes approached significance, $\chi^2(1) = 3.61$, $p = 0.06$. Furthermore, the number of thought/feeling entries that was congruent with the targets' direction of gaze was similar in both videotapes (i.e. 35.7%).

Empathic Accuracy Task: Empathic Accuracy Measures

Two measures of empathic accuracy were computed from comparisons of the actual and inferred thought/feeling entries. The first measure, *valence accuracy*, is an index of the proportion of instances in which each inference about the overall emotional tone (negative, neutral or positive) matches the actual valence label assigned to each entry. Following the logic and procedure described by Ickes and colleagues (Ickes et al. 1990b), we computed the number of matches and mismatches between the rated valence of the actual entries and the corresponding inferences for each perceiver. We then computed the percentage of correct matches for each perceiver.

The second measure, *content accuracy*, is an index of the degree to which a perceiver's written description of the inferred content matches the target's actual content of the sentence. As suggested by Ickes et al. (1990a), five naïve and independent judges were instructed to compare each perceiver's inferred thought/feeling entry with the corresponding original thought/feeling entry and to rate the level of similarity on a three-point scale, ranging from 0 (essentially different content) through 1 (somewhat similar but not the same content) to 2 (essentially the same content), with "I don't know" and missing responses rated 0. The internal consistency (Cronbach's alpha) of the five judges' content accuracy ratings was 0.88 for the ASD group and 0.89 for the control group. The mean of the accuracy scores assigned by the five judges was calculated for each individual inference. We then calculated aggregated accuracy scores from the averaged scores for each inference. The mean ratings were therefore summed across the thought/feeling entries for each of the two videotapes. These summed scores were divided by the total number of thought/feeling entries for each target and multiplied by 100. By doing so, the minimum and maximum empathic accuracy scores of each perceiver could vary between 0 (total inaccuracy) and 100 (total accuracy).

Following the logic and procedures described by Ickes and colleagues (Ickes et al. 1990a, b), a measure of baseline empathic accuracy was similarly derived for each of the perceivers. This logic suggested that perceivers could simply make a fortuitous guess about the content of the thought/feeling entry and be sometimes correct because of chance alone (Ickes et al. 1990b). A straightforward method to estimate the baseline accuracy level that reflects this false accuracy component is to randomly pair each set of the actual thought/feeling entries with the corresponding set of the perceiver's inferences and to ask independent judges to make similarity ratings of these randomly paired actual/inferred entries. The internal consistency of the baseline accuracy provided by the same five judges was 0.69 for the ASD group and 0.59 for the control group. The mean of the baseline empathic accuracy scores rated by the five judges was further calculated for each individual inference and aggregated accuracy scores were calculated from the averaged scores for each inference. We then summed the mean ratings and these values were divided by the total amount of thought/feeling entries and multiplied by 100. By subtracting for each perceiver the baseline empathic accuracy scores from the original content scores, we derived a revised measure of global empathic accuracy score for each perceiver.

Following this logic, we calculated some more refined empathic accuracy scores for each videotape. First, the empathic accuracy scores were calculated for all thoughts and feelings belonging to (a) the most easy-to-infer thought/feeling entries, (b) the most difficult-to-infer thought/feeling entries and (c) the remaining moderate difficult-to-infer thought/feeling entries. Second, the empathic accuracy scores for each videotape were calculated for all thought/feeling entries belonging to (a) the most abstract thought/feeling entries, (b) the most concrete thought/feeling entries and (c) the moderate abstract thoughts and feelings. Third, we calculated the empathic accuracy scores for all thought/feeling entries judged by the targets as (a) negative, (b) neutral and (c) positive in affective tone. Finally, we computed for each videotape the empathic accuracy scores for all the thought/feeling entries with the content of the thought/feeling entries whether or not to be congruent with (a) the targets' verbal communication and (b) with the targets' direction of gaze.

Results

Empathic Accuracy: Content Accuracy

Although the videotapes were presented in a counterbalanced order, a preliminary analysis was conducted to reveal whether there were any order effects in the empathic accuracy scores. A MANOVA with Group (ASD versus Control) and Order of videotape (video 1 versus video 2) as factors and the empathic accuracy scores on the videotapes as dependent variables was conducted. Analysis revealed no significant main effect of Order, with $F(2,39) = 2.32$, $p = 0.11$, and no significant interaction effect between Order of videotape and Group, with $F(2,39) < 1$.

A 2 (Group: ASD versus Control) × 2 (Structure: Video 1 and Video 2) ANOVA was conducted, with Structure as within-subject factor and Group as between-subject factor. The analysis revealed a significant main effect of Group, in favour of the control subjects, $F(1,42) = 7.70$, $p < 0.01$. No significant main effect of Structure was found, $F(1,42) = 1.40$. As expected, a significant interaction effect between Group and Structure revealed greater difficulties in inferring the thoughts and feelings of the targets in the unstructured situation than from the structured one among the ASD group, whereas the control group showed no difference, $F(1,42) = 6.61$, $p = 0.01$.

An ANOVA testing for between-group differences confirmed that the subjects with ASD were found to make less accurate inferences of the thoughts and feelings of the targets in Video 1, $F(1,42) = 14.46$, $p < 0.0001$. The mean content accuracy of the subjects with ASD was 15.81% (SD = 10.18) and the mean content accuracy of the control subjects was 31.68% (SD = 16.73). No significant between-group difference was found on video 2, $F(1,42) < 1$. In video 2, the mean content accuracy of the ASD group was 24.61% (SD = 15.64) and the mean content accuracy of the control group was 28.44% (SD = 12.94).[3]

A comparison between the scores for video 1 and video 2 in each group revealed that the subjects with ASD performed better on the structured video than on the less structured one, $t(21) = -2.95$, $p < 0.01$, and that there was no difference in the control group, $t(21) = 0.90$. We further analysed the more refined content accuracy scores.

Difficulty

The mean content accuracy scores of the thought/feeling entries belonging to the most easy-to-infer thought/feeling entries, the most difficult-to-infer thought/feeling entries and the moderately difficult-to-infer thought/feeling entries are shown in Table 2, represented separately for both videotapes and for each group. A 2 (Group: ASD versus Control) × 2 (Structure: Video 1 and Video 2) × 3 (Difficulty: Easy, Moderate and Difficult) ANOVA was conducted on the empathic accuracy scores, with Group as between-subject factor and Structure and Difficulty as within-subject factors (see Table 3). A significant main effect for Difficulty was found. Follow-up analyses revealed that the most easy-to-infer thought/feeling entries were more accurately inferred than the moderate difficult-to-infer thought/feeling entries, $F(1,42) = 20.85$, $p < 0.0001$, and that the moderate difficult-to-infer thoughts and feelings were more accurately inferred than the most difficult-to-infer thoughts and feelings, $F(1,42) = 33.92$, $p < 0.0001$. As shown in Table 3, the interaction between Group and Difficulty yielded not significant, indicating that both groups have the same tendency to be less accurate in inferring more difficult thought/feeling entries. However, the interaction between Structure and Difficulty was significant. Paired *t*-tests suggest that the moderate difficult-to-infer thought/feeling entries of the structured videotape (video 2) were better inferred than those of the unstructured videotape, with $t(43) = -3.01$, $p < 0.005$, whereas no difference were found with regard to the easy-to-infer and the difficult-to-infer thoughts and feelings, with $t(43) = 1.15$ and $t(43) = -0.59$, respectively. The three-way interaction (Group × Structure × Difficulty) yielded not significant. Further comparisons between the empathic accuracy scores for the structured and unstructured videotape in each group (paired *t*-tests) revealed that subjects with ASD were significantly better at inferring the moderate difficult-to-infer

[3] A non-parametric test (Mann–Whitney *U*) revealed similar results, with $z = -3.35$, $p \leq 0.001$ for video 1 and $z = -1.66$, ns, for video 2. The mean rank for the controls was 28.98 on video 1 and 25.70 on video 2, while the mean rank for the ASD group was 16.02 on video 1 and 19.30 on video 2.

Table 2 Content empathic accuracy along the nature of the thought/feeling entries (in percentages)

	ASD subjects					Control subjects					All subjects			
	Video 1: Unstructured		Video 2: Structured		$t(21)$	Video 1: Unstructured		Video 2: Structured		$t(21)$	Video 1: Unstructured		Video 2: Structured	
	M	SD	M	SD		M	SD	M	SD		M	SD	M	SD
Difficulty														
Easy	30.65	20.30	37.01	29.72	–0.99	53.12	27.58	36.62	18.36	3.21[b]	41.88	26.49	36.82	24.42
Moderate	13.38	14.08	25.66	16.09	–3.32[b]	28.77	20.15	32.80	14.67	–1.04	21.07	18.86	29.23	15.63
Difficult	5.84	14.47	12.05	20.10	–1.13	16.10	15.81	12.20	18.52	0.39	10.97	15.85	13.12	19.14
Abstractness														
Concrete	23.94	21.17	39.22	26.44	–2.88[a]	35.15	29.72	35.06	20.61	0.02	29.54	26.12	37.14	23.52
Moderate	15.52	16.04	22.44	16.02	–1.81*t*	34.34	20.58	30.63	12.18	0.78	24.93	20.56	26.53	14.66
Abstract	10.81	13.58	5.45	12.08	1.2	25.56	24.96	12.00	23.55	2.03*t*	18.18	21.21	8.72	19.02
Affective Tone														
Negative	16.57	10.48	22.93	17.35	–1.43	44.04	23.78	32.02	27.01	2.00	30.30	22.87	27.47	22.90
Neutral	20.61	16.13	21.59	18.22	–0.27	26.57	18.23	30.57	17.13	–0.72	23.59	17.27	26.08	18.05
Positive	10.82	17.85	28.18	22.78	–3.51[b]	25.18	16.79	23.97	13.47	0.26	18.00	18.60	26.07	18.62
Verbalisation														
Congruent	23.30	20.14	32.61	17.66	–2.24[a]	41.02	24.47	36.00	12.27	1.12	32.16	23.89	34.30	15.12
Non-congruent	18.91	9.54	20.35	16.63	–0.38	34.55	14.79	26.01	13.86	1.99*t*	26.73	14.62	23.18	15.40
Gazing														
Congruent	21.27	13.01	32.73	16.44	–3.08[b]	42.36	18.49	36.45	14.27	1.43	31.82	19.07	34.59	15.33
Non-congruent	19.55	11.70	23.69	16.80	–1.18	33.08	16.76	28.54	12.58	1.34	26.31	15.84	26.11	14.87

[a] < 0.05

[b] < 0.01

$0.05 < t < 0.10$

Table 3 Repeated measures along the topics

Topic	*df*	*F*-values			
		Main effect	Group × topic	Structure × topic	Group × Structure × topic
Difficulty	(2,41)	68.74***	<1	3.79*	1.90
Abstractness	(2,41)	20.68***	1.38	6.62*	<1
Affective tone	(2,41)	4.40*	4.81**	3.63*	3.38*
Congruent with					
Verbal communication	(2,41)	19.10***	<1	1.92	<1
Direction of gaze	(2,41)	15.32***	<1	<1	1.83

***$p < 0.001$
**$p < 0.01$
*$p < 0.05$

thoughts and feelings of the structured videotape than those of the unstructured videotape, whereas the control subjects were significantly better at inferring the most easy-to-infer thoughts and feelings of the unstructured videotape than those of the structured videotape (see Table 2).

Abstractness

Similarly, Table 2 presents for each group the mean content accuracy scores of the thoughts and feelings belonging to one of the three categories of abstractness. A 2 (Group: ASD versus Control) × 2 (Structure: Video 1 and Video 2) × 3 (Abstractness: Concrete, Moderate and Abstract) ANOVA was conducted on the empathic accuracy scores, with Group as between-subject factor and Structure and Abstractness as within-subject factors (see Table 3). A significant main effect for Abstractness was found. Follow-up analyses revealed that the most concrete thought/feeling entries were more accurately inferred than the moderate abstract thought/feeling entries, $F(1,42) = 6.52$, $p = 0.01$, and that the moderate abstract thoughts and feelings were more accurately inferred than the most abstract thoughts and feelings, $F(1,42) = 26.83$, $p < 0.0001$. As shown in Table 3, the interaction between Group and Abstractness was not significant, indicating that both groups have the same tendency to be less accurate in inferring more abstract thought/feeling entries. However, the interaction between Structure and Abstractness was significant. Paired *t*-tests suggest that the most concrete thought/feeling entries of the structured videotape were slightly better inferred than those of the unstructured videotape, $t(43) = -1.96$, $p = 0.06$. Furthermore, the most abstract thought/feeling entries of the unstructured videotape were better inferred than those of the structured videotape, $t(43) = 2.35$, $p < 0.05$. No difference were found with regard to the moderate abstract thoughts and feelings, with $t(43) = -0.51$. The Group × Structure × Difficulty interaction yielded not significant.

In addition, comparisons between the empathic accuracy scores for the structured and unstructured videotape in each group revealed that subjects with ASD were significantly better at inferring the concrete thoughts and feelings of the structured videotape than those of the unstructured videotape and slightly better at inferring the moderate abstract thoughts and feelings of the structured videotape than those of the unstructured videotape, whereas the control subjects were slightly better at inferring the most abstract-to-infer thoughts and feelings of the unstructured videotape than those of the structured videotape (see Table 2).

Affective Tone

The mean content accuracy scores of the thought/feeling entries judged by the targets as negative, neutral and positive in affective tone are shown in Table 2, presented separately for both videotapes and for each group. A 2 (Group: ASD versus Control) × 2 (Structure: Video 1 and Video 2) × 3 (Affective Tone: Negative, Neutral and Positive) ANOVA was conducted on the empathic accuracy scores, with Group as between-subject factor and Structure and Affective Tone as within-subject factors (see Table 3). A significant main effect of Affective Tone was found. Follow-up analyses indicated that negative thought/feeling entries were slightly better inferred than neutral thought/feeling entries, with $F(1,42) = 3.38$, $p = 0.07$, and that negative thoughts and feelings were significantly better inferred than positive thoughts and feelings $F(1,42) = 8.61$, $p < 0.01$. No differences were found between the neutral and positive thoughts and feelings, with $F(1,42) = 1.33$. As shown in Table 3, the interaction between Group × Affective Tone yielded significant. Paired *t*-tests suggest that the control group inferred thoughts and feelings with a negative tone significantly better than thoughts and feelings with a neutral tone, $t(21) = 2.52$, $p < 0.05$, and that they inferred thoughts and feelings with a neutral tone slightly better than thoughts and feelings with a positive tone, $t(21) = 1.90$, $p = 0.07$, so that the difference between the empathic accuracy belonging to thought/feeling entries with a negative versus positive tone was significant, $t(21) = 3.67$, $p < 0.005$. Although subjects with ASD have somewhat higher empathic accuracy scores for thoughts and feelings with a neutral tone (see Table 2), paired *t*-tests revealed no significant differences in inferring the thoughts and feelings with different affective tones. A series of between-group analyses (ANOVA) for each videotape and Affective Tone revealed that the control subjects were significantly better than the ASD subjects at inferring the negative and positive thoughts and feelings of the unstructured videotape, with $F(1,42) = 24.58$, $p < 0.0001$ and $F(1,42) = 7.56$, $p < 0.01$. No other significant between-group differences were found.

Furthermore, the interaction between Structure × Affective tone was significant (see Table 3). Paired *t*-tests suggest that the thought/feeling entries with a positive tone of the structured videotape were significant better inferred than those of the unstructured videotape, $t(43) = -2.19$, $p < 0.05$. No differences were found with regard to the negative and neutral thoughts and feelings, with $t(43) = 0.72$ an $t(43) = -0.76$, respectively. Finally, the Group × Structure × Affective Tone interaction yielded significant, indicating that subjects with autism performed much better in the structured condition for the negative and the positive entries, whereas this pattern was not so clear in the control subjects.

When we compared the empathic accuracy scores for the structured and unstructured videotape in each group (paired *t*-tests), subjects with ASD were found to be significantly better at inferring the positive thoughts and feelings of the structured videotape than those of the unstructured videotape, whereas no differences were found for the control subjects (see Table 2).

Congruence

The mean content accuracy scores of thought/feeling entries along the congruence with the targets' verbal communication and the targets' gazes are shown in Table 2, represented separately for both videotapes and each group. We conducted two 2 (Group: ASD versus Control) × 2 (Structure: Video 1 versus Video 2) × 2 (Congruence: Congruent versus Non-Congruent) repeated measures (ANOVA), with Group as between-subject factor and Structure and Congruence as within-subjects factors (see Table 3). The analyses revealed two significant main effects of Congruence, indicating that the subjects reached higher accuracy rates for thoughts and feelings that are congruent with the targets' verbal communication and the targets' direction of gaze than those that are not congruent with the targets' communication and direction of gaze. No significant interaction effects (Congruence × Group) were found in both analyses, but the main effects of Group and Congruence were found to be additive ($p > 0.25$) for the verbalisation. The additive effect suggests that the lowest empathic accuracy among all other empathic accuracy rates was obtained by the ASD group inferring thoughts and feelings that are not congruent with the targets' communication. Furthermore, the Structure × Congruence interaction with regard to the targets' verbal communication and the targets' gazes were not significant. In addition, the three-way interaction with regard to the targets' verbal communication and the targets' gazes yielded not significant.

As shown in Table 2, comparisons between the empathic accuracy scores for the structured and unstructured videotape in each group revealed that subjects with ASD were significantly better at inferring the thoughts and feelings that are congruent with the targets' verbal communication and direction of gaze of the structured videotape than those of the unstructured videotape, whereas the control subjects were slightly better at inferring the thoughts and feelings of the unstructured videotape that are non-congruent with the targets' communication than those of the structured videotape.

Empathic Accuracy: Valence Accuracy

The mean valence accuracy of the subjects with ASD was 35.39% (SD = 9.60) for video 1 and 35.71% (SD = 6.70) for video 2. The mean valence accuracy of the control subjects was 35.55% (SD = 10.88) for video 1 and 40.26% (SD = 8.54) for video 2. In order to exclude the possibility that the subjects made a lucky guess about the affective tone of the thoughts and feelings and were correct on a few rare occasions because of chance alone (Ickes et al. 1990b), we conducted for each videotape and for each group a one-sample *t*-test with 33.33% as test value. In the first videotape, the valence accuracy of both groups did not significantly differ from 33.33%, with $t(21) = 1.17$ for the ASD group and $t(21) = 1.10$ for the control group, indicating that the valence accuracy scores were not significantly different from coincidence. In the second videotape, a significant difference was found for the control group, $t(21) = 4.24$, $p < 0.0001$, and a trend was found for the group with ASD, $t(21) = 1.90$, $p = 0.07$. An ANOVA with the valence accuracy score of video 2 as dependent variable and group as factor, revealed a significant between-group difference in favour of the control subjects, $F(1, 42) = 4.15$, $p < 0.05$.

Correlates of the Empathic Accuracy Task

The correlations and intercorrelations between the content empathic accuracy scores, the valence accuracy scores and the total IQ scores were calculated beyond both groups. A significant positive correlation (Pearson) was found between the content accuracy score on video 1 and that on video 2, $r(44) = 0.40$, $p < 0.01$. Furthermore, we found a significant positive correlation between the subjects' total IQ scores and the content accuracy scores on video 1, $r(44) = 0.38$, $p = 0.01$, and video 2, $r(44) = 0.32$, $p < 0.05$. No other significant correlations were found.

Furthermore, the correlations and intercorrelations between the content empathic accuracy scores, the valence accuracy scores and the total IQ scores were calculated for both groups separately. The correlation (Pearson) between the content accuracy score on video 1 and that on video 2 approached significance for the control subjects, $r(22) = 0.37$, $p = 0.09$, and was significant for the subjects with ASD, $r(22) = 0.48$, $p < 0.05$. Furthermore, we found a significant positive correlation between the total IQ scores of the control adults and their content accuracy scores on video 1, $r(22) = 0.60$, $p < 0.01$, and between the total IQ scores of the control adults and their valence accuracy scores on video 1, $r(22) = 0.50$, $p < 0.05$. No other significant correlations were found.

We further used the formula[4] of Hays (1994) to transform the Pearson correlations into z-scores. The transformations revealed that the strength of the associations were equally strong in both groups, with the exception of the significant association between Total IQ and valence accuracy on Video 1 ($z = 2.24$, $p < 0.05$).

Discussion

This study investigated the empathic accuracy of subjects with ASD when confronted with two conversations between two strangers: a highly structured conversation and a less structured or more naturalistic one. The results suggest that structure does matter for the mind-reading abilities of subjects with ASD, in that way that mind-reading differences between young adults with ASD and typically developing controls were more pronounced when subjects had to infer thoughts and feelings of other persons in the less structured conversation.

[4] $\frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$.

Since the mind-reading impairments of the subjects with ASD were only apparent in the less structured conversation, we tried to understand more fully these findings by studying empathic accuracy along variables such as difficulty, abstractness and affective tone of the videotaped strangers' thoughts and feelings, the videotaped persons' verbalisation and their direction of gaze. These more refined analyses suggest that young adults with ASD use, at least to some extent, the same strategies as typically developing controls to infer other people's thoughts and feelings. Support for this comes from the absence of any interaction effect between group and the variables discussed, with the exception of the Group and Affective Tone interaction effect and the Group and Structure and Affective Tone interaction effect. The first interaction effect suggests that adults with ASD were slightly better in inferring when the affective content of the conversation was neutral, whereas control persons were better able to infer when the target individuals expressed negative affect. The latter interaction effect suggest that adults with ASD benefit more from the structured situation when they have to infer positive and—to a lesser degree—negative thoughts and feelings, whereas this pattern is not so clear in the control adults. One possibility for this is that the more disturbing affective cues might help clarify others' inner states for typically developing subjects, whereas subjects with ASD may find expressions of negative or positive affect in a high naturalistic conversation more distracting than helpful. Yet, several main effects were indicating that in both videotapes the empathic accuracy scores of all subjects decreased with an increasing level of difficulty. The analyses revealed that the most concrete thoughts and feelings were better inferred than the most abstract thoughts and feelings. These results agree with the notion that subjects with ASD have greater difficulty with symbolic or abstract language concepts than with concrete facts and descriptions (Mesibov and Shea 2001). However, it should be noted that this is also the case for the control group. With regard to congruence, in both videotapes it was found that thoughts and feelings that are congruent with the targets' communication or the targets' eye gazes are better inferred than incongruent thoughts and feelings, which is -in a way- consistent with other findings (Golan et al. 2006; Grossman et al. 2000). Words and eye gazes are just two ways to show our intentions. Other non-verbal actions (e.g. body movements, smiles) can speak as loud as words because they too can act as clues to our intended meaning (Happé 1994b). The fact that the inference ability of both groups was better when the verbalisations and eye gazes supported the targets' covert thoughts and feelings, suggests that young adults with ASD are able to recognise, at least to some part, the supportive character of these variables. None of the above-mentioned variables are thus able to explain the differences in content empathic accuracy between adults with ASD and control adults in the more chaotic conversation.

With regard to the valence accuracy, we first calculated whether or not the scores of the participants differed from chance. The analyses revealed that the valence accuracy scores of both groups were only significantly different from chance in the more structured videotape. In this videotape, the control subjects were significantly better than the subjects with ASD in inferring the emotional tone of target's thoughts and feelings. One possibility is that subjects have more difficulties in the more chaotic conversation with inferring the affective tone of other persons' thoughts and feelings, which might have given them cause to guess.

In sum, this study demonstrated the significance of structure to the empathic accuracy of young adults with ASD and showed that the empathic accuracy design offers opportunities to test more refined hypotheses with regard to the "on line" mind-reading abilities of subjects with ASD. Furthermore, the findings might explain why some high-functioning subjects with ASD are able to pass through traditional mind-reading tasks, although they have severe social-functioning problems in daily life. While laboratory tasks provide clear structure, daily life situations are characterised by circumstances beyond our control. One implication is that, if mind-reading difficulties of subjects with ASD become mainly apparent in less structured situations, then it seems more appropriate to develop mind-reading tasks in which the structure is reduced, so that even more subtle mind-reading impairments can be detected.

It should be noted that the mind-reading performance of normally intelligent individuals with ASD is subject to disparity in age (Happé 1995; Yirmiya et al. 1996). Several studies have demonstrated that the majority of children with ASD fail to pass traditional as well as more advanced mind-reading tasks (Baron-Cohen et al. 2000). However, the greater part of high-functioning adults with ASD seems to have no difficulties with traditional mind-reading tasks. Their mind-reading impairments become mainly apparent with the use of more advanced or naturalistic mind-reading tasks (Kleinman et al. 2001; Ponnet et al. 2004; Roeyers et al. 2001). Therefore, it can be argued that adults with ASD have more subtle mind-reading impairments than young children with ASD. It is unclear whether adults with ASD have developed compensatory strategies to camouflage their mind-reading impairments or whether their mind-reading skills improved spontaneously with growing age. One way to find out is to administer traditional and more naturalistic mind-reading tasks from teenagers with ASD. It has been suggested that teenagers with ASD have no problems with traditional mind-reading tasks (Dahlgren and Trillingsgaard 1996). However, since little research has been done with this population, it has to be seen how they perform on more naturalistic mind-reading tasks. One might suggest that the use of naturalistic tasks would clarify even more the mind-reading impairments in teenagers with ASD than those in adults with ASD.

A limitation of the present study is that the Empathic Accuracy Task cannot be considered as a highly naturalistic real-life task, at best the task is a highly dynamic one. Both videotapes were manipulated so that one of the tapes was more structured than the other. This manipulation was needed to test the hypothesis, but also restricts the naturalistic character of the task. Furthermore, inherent to the methodology of the empathic accuracy design, subjects are permitted to use as much time as needed to infer the thoughts and feelings of other persons. However, in real life, thoughts and feelings have to be inferred "hic et nunc". In a way, this implies that the design does not mirror normal life. Another possible issue of concern that applies to this study is the criteria that were used to recruit our ASD sample. Since we were interested in between-group differences and not in within-group differences, we did not distinguish subjects who received an early diagnosis from those who received a late diagnosis. Similarly, we made no distinction between subjects with Asperger syndrome from those with autism or PDD-NOS. However, previous studies already revealed that the empathic accuracy design is able to capture mind-reading impairments of adults with autism and PDD-NOS (Roeyers et al. 2001) as well as adults with Asperger syndrome (Ponnet et al. 2004). Based on these results and considering the diagnostic difficulties that clinicians experience to distinguish autism from Asperger syndrome, or the controversy whether or not autism and Asperger syndrome lie on the same continuum (Bishop 2000; Mayes and Calhoun 2001; Wing 1998), the ASD group was taken as a whole. Besides, the sample size of the ASD group (n = 22) is not large enough to reveal any within-group differences. However, this might be an interesting issue to explore in future studies with the empathic accuracy design and an enlarged clinical sample. It seems plausible that the empathic accuracy ability of subjects with ASD is related to the severity of their social-communication disabilities. It would therefore be interesting to investigate to what extent -for instance- adults with Asperger syndrome differ from adults with autism. Furthermore, one may speculate whether adults who received a late diagnosis perform better than those who received an early diagnosis.

The question remains which variables can account for the empathic accuracy differences between subjects with ASD and typically developing subjects. The correlations suggest that the various indices of the content empathic accuracy are, with a single exception, independent of each other. The observed correlation suggests that only 16% of variability in performance is shared across both videotape tasks. What this pattern of results might suggest is that performance on both videotapes is not related to the ability of individuals: Even within the control condition, subjects who achieve higher scores on one videotape are not that likely to achieve higher scores on the other videotape. Therefore, one might argue that achievement is largely task-dependent, and that characteristics other than difficulty, abstractness, congruence and so on limit the performance on one videotape and affect performance differently on the other videotape. However, it should be noted that even studies with typically developing individuals still failed to identify reliable and replicable individual-difference correlates of "good" versus "poor" perceivers (Ickes et al. 2000). One of the problems in the present study is that we do not know to what specific condition we are able to attribute the fact that there is minimal evidence of any one individual differences dimension being measured by both videotapes. In a way, this means that we are not able to attribute the between-group differences in performance to group differences in some ability that affects performance on both tapes.

In future empathic accuracy studies, it might be interesting to sample more interactions, so that raters have to judge a larger number of targets, and to analyse the relative contribution of perceiver, target and relationship effects to the empathic accuracy scores, for instance using the Social Relations Model (see Kenny 1994). In addition, it might be interesting to explore the significance of subjects' with ASD personal variables to their mind-reading performance. It seems plausible that the mind-reading performance of subjects with ASD is related to variables such as the severity of subjects' social-communication disabilities, their motivation, or the subjects' perception of their own mind-reading abilities. Therefore, the use of multiple instruments at the same time may contribute towards a better understanding of the mind-reading impairments in subjects with ASD.

Acknowledgment This study was supported by a research grant (3.G.009.000) of the Fund for Scientific Research, Flanders, Belgium. We would like to thank our undergraduate research assistants Herlinde Van de Velde, Sandra Van Hoebrouck, Veerle Poppe and Maaike De Clerck for help in data collection. We are indebted to the adults and their parents, who participated in the present study.

Appendix A

The eight-point list to structure video 2

1. What's your name?
2. Where do you live?
3. How do you come to be here?
4. Where do you work or study?
5. What are your leisure activities?
6. How old are you?
7. Do you have specific hobbies?
8. What are our common interests?

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

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

Baron-Cohen S. (2003). *The essential difference: The truth about the male and female brain*. NY: Basic Books.

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

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, J. (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience*. Oxford: University Press.

Baron-Cohen, S., Wheelwright, 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*, 241–253.

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.

Bauminger, N., & Shulman, C. (2003). The development and maintenance of friendship in high-functioning children with autism. Maternal perceptions. *Autism, 7*, 81–97.

Bishop, D. (2000). What's so special about Asperger syndrome ? The need for further exploration of the borderlands of autism. In A. Klin, F. Volkmar, & S. Sparrow (Eds.), *Asperger syndrome* (pp. 254–277). New York: Guilford Press.

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

Boucher, J. (1989). The theory of mind hypothesis of autism: Explanation, evidence and assessment. *British Journal of Disorders of Communication, 24*, 181–198.

Buysse, A., & Ickes, W. (1999). Topic-relevant cognition and empathic accuracy in laboratory discussions of safer sex. *Psychology and Health, 14*, 351–366.

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

De Clercq, A., Buysse, A., Roeyers, H., Ickes, W., Ponnet, K., & Verhofstadt, L. (2001). VIDANN: A video annotation system. *Behavior Research Methods, Instruments, and Computers, 33*, 159–166.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J., Wolf, T., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*, 623–636.

Eisenberg, N., Murphy, B., & Shepard, S. (1997). The development of empathic accuracy. In Ickes, W. (Ed.), *Empathic accuracy* (pp. 73–116). New York: Guilford Press.

Eisenberg, N., Shea, C. L., Carlo, G., & Knight, G. (1991). Empathy-related responding and cognition: A "chicken and the egg" dilemma. In W. Kurtines, & J. Gewirtz (Eds.), *Handbook of moral behavior and development: Vol. 2. Research* (pp. 63–68). Hillsdale NJ: Lawrence Erlbaum.

Gillberg, C., Nordin, V., & Ehlers, S. (1996). Early detection of autism. Diagnostic instrument for clinicians. *European Child and Adolescent Psychiatry, 5*, 67–74.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders, 36*, 169–183.

Grossman, J., Klin, A., Carter, A., & Volkmar, F. (2000). Verbal bias in the recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*, 369–379.

Happé, F. (1994a). 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. (1994b). *Autism: An introduction to psychological theory*. London: University College London Press.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*, 843–855.

Hays, W. L. (1994). *Statistics*. Fifth edition. Orlando: Harcourt Brace College Publishers.

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.

Howlin, P. (1997). *Autism: Preparing for Adulthood*. New York: Routledge.

Ickes, W. (1993). Empathic Accuracy. *Journal of Personality, 61*, 587–610.

Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. L. (1990a). Implementing and using the dyadic interaction paradigm. In C. Hendrick, & M. S. Clark (Eds.), *Review of personality and social psychology: Research methods in personality and social psychology* (Vol. 11, pp. 16–44). Newbury Park, CA: Sage.

Ickes, W., Buysse, A., Pham, H., Rivers, K., Erickson, J., Hancock, M., Kelleher, J., & Gesn, P. (2000). On the difficulty of distinguishing 'good' and 'poor' perceivers: A social relations analysis of empathic accuracy data. *Personal Relationships, 7*, 219–234.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990b). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59*, 730–742.

Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment and validation. *Journal of Personality and Social Psychology, 51*, 66–82.

Kenny, D. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

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

Klin, A., & Jones, W. (2006). Attributing social and physical meaning to ambiguous visual displays in individuals with higher-functioning autism spectrum disorders. *Brain and Cognition, 61*, 40–53.

Klin, A, Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*, 809–816.

Lord, C. (1993) The complexity of social behaviour in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds* (pp. 292–316). New York: Oxford University Press.

Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology, 68*, 854–869.

Mayes, S., & Calhoun, S. (2001). Non-significance of early speech delay in children with autism and normal intelligence and implications for DSM-IV Asperger's disorder. *Autism, 5*, 81–94.

Mesibov, G. (1992). Treatment issues with high-functioning adolescents and adults with autism. In E. Schopler, & G. Mesibov (Eds.), *High-functioning individuals with autism* (pp. 143–156). New York: Plenum Press.

Mesibov, G., & Shea, V. (2001). *Understanding Asperger syndrome and high-functioning autism*. Netherlands: Kluwer Academic Publishers.

Perner, J., & Wimmer, H. (1985). "John thinks tha Mary thinks that..." Attribution of second order beliefs by 5–10 year old children. *Journal of Experimental Child Psychology, 39*, 437–471.

Ponnet, K., Roeyers, H., Buysse, A., De Clercq, A., & Van der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism, 8*, 249–266.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioural and Brain Sciences, 4*, 515–526.

Prior, M. R., Dahlstrom, B., & Squires, T. L. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry, 31*, 587–601.

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychology and Psychiatry, 42*, 271–278.

Rutterford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders, 32*, 189–194.

Simpson, J. A., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationship. *Journal of Personality and Social Psychology, 69*, 629–641.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33*, 591–605.

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology, 62*, 787–797.

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics, 16*, 241–256.

Wechsler, D. (2000). *WAIS-III: Wechsler Adult Intelligence Test* (3rd ed.). Lisse: Swets & Zeitlinger.

Yirmiya, N., Solomonica-Levi, D., Shulman, C., & Pilowsky, T. (1996). Theory of mind abilities in individuals with autism, down syndrome, and mental retardation of unknown etiology: The role of age and intelligence. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*, 1003–1014.

Zager, D. (1999). *Autism: Identification, education, and treatment* (2nd ed.). New York: Lawrence Erlbaum Associates.