J Autism Dev Disord (2006) 36:623–636
DOI 10.1007/s10803-006-0107-0

ORIGINAL PAPER

Introducing MASC: A Movie for the Assessment of Social Cognition

Isabel Dziobek · Stefan Fleck · Elke Kalbe · Kimberley Rogers · Jason Hassenstab · Matthias Brand · Josef Kessler · Jan K. Woike · Oliver T. Wolf · Antonio Convit

Published online: 6 June 2006

Abstract In the present study we introduce a sensitive video-based test for the evaluation of subtle mindreading difficulties: the Movie for the Assessment of Social Cognition (MASC). This new mindreading tool involves watching a short film and answering questions referring to the actors' mental states. A group of adults with Asperger syndrome ($n = 19$) and well-matched control subjects ($n = 20$) were administered the MASC and three other mindreading tools as part of a broader neuropsychological testing session. Compared to control subjects, Asperger individuals exhibited marked and selective difficulties in social cognition. A Receiver Operating Characteristic (ROC) analysis for the mindreading tests identified the MASC as discriminating the diagnostic groups most accurately. Issues pertaining to the multidimensionality of the social cognition construct are discussed.

Keywords Asperger syndrome · Theory of Mind · Mindreading · Naturalistic test formats · Emotion recognition

I. Dziobek (✉) · J. Hassenstab · A. Convit
Center for Brain Health, New York University School of Medicine, New York, NY, USA
e-mail: dziobi01@med.nyu.edu

E. Kalbe · J. Kessler
Department of Neurology, University Clinic Cologne, Cologne, Germany

O. T. Wolf · I. Dziobek
Department of Psychology, University of Bielefeld, Bielefeld, Germany

S. Fleck
Department of Neuroscience and Rehabilitation, University of Cologne, Cologne, Germany

M. Brand
Department of Physiological Psychology, University of Bielefeld, Bielefeld, Germany

K. Rogers
Department of Psychology at the Graduate Center of City University of New York (CUNY), New York, NY, USA

J. K. Woike
Department of Psychological Methodology, Ruhr-University of Bochum, Bochum, Germany

The ability to attribute mental states to oneself and others is referred to as social cognition or theory of mind. Making social cognitive inferences is crucial for successful social interactions because they mediate an understanding of the dispositions and intentions of others and lead to the correct prediction of behavior (Brothers, 1990). Individuals diagnosed with the neurodevelopmental disorder Asperger syndrome (AS) have core, often selective, deficits in inferring others' mental states (Baron-Cohen, 1995; Jolliffe & Baron-Cohen, 1999). In the DSM-IV, Asperger syndrome is listed among the pervasive developmental disorders (PDD) and characterized as a condition with impairments in social interactions and the presence of restricted interests and behaviors (American Psychiatric Association, 1994). In accordance with Hans Asperger (1944), who first described the clinical picture, a diagnosis of AS requires the patient to not display a general delay in language or cognitive development (American Psychiatric Association, 1994), representing the essential difference to autism. Despite persisting debates as to whether Asperger syndrome is qualitatively different from a more high-functioning form of autism (HFA; Howlin, 2003; Klin & Volkmar, 1997), most authors regard Asperger syndrome as belonging to the autism spectrum, based on the large amount of shared diagnostic features (Volkmar, Klin,

Schultz, Rubin, & Bronen, 2000; Wing, 1997). Consequently, we also included, when relevant, work in autism spectrum disorders. When referring to individuals on the autism spectrum in this manuscript, we have purposely avoided ''person first'' language (i.e. individuals *with* autism) and instead used the term autism descriptively (i.e. autistic individuals) in order to reflect language preferences within the autism community.

The social cognitive deficits seen in autism spectrum disorders have received considerable attention over the last decade, since they are likely a key contributor to the broad social impairments observed in affected individuals. Early research in autistic children has used ''first-order false-belief'' tasks, which involve an understanding that others may have different mental states from themselves and may therefore hold a false belief (Wimmer & Perner, 1983). While unaffected children pass first- and second-order tasks by the age of three and six respectively, numerous studies have demonstrated that autistic children have difficulties in shifting their perspective to judge other peoples' simple mental states (Baron-Cohen, Leslie, & Frith, 1985; Pilowsky, Yirmiya, Arbelle, & Mozes, 2000; Reed & Peterson, 1990). A variety of other tasks, such as tests of deception (Baron-Cohen, 1992), recognition of faux pas (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999), or comprehension of intentions in communication (Happé, 1993), have demonstrated similar mindreading difficulties.

For higher functioning autistic adults, however, most of these tests do not pose a big enough challenge. For example, several authors have reported that their patient groups succeeded on first- and second-order false belief tasks (Bowler, 1992; Happé, 1994; Jolliffe & Baron-Cohen, 1999; Ozonoff, Rogers, & Pennington, 1991). This is in contrast with clear problems in social cognition apparent in everyday life and has led to the development of more advanced tests in recent years.

A first step was undertaken by Happé (1994), who assessed story comprehension in a group of higher functioning autistic adults. Her ''Strange Stories Task'' requires subjects to make inferences about the mental states of story characters, using concepts such as double bluff, mistakes, irony, or white lie. This initial study found clear group effects, but also a relationship between verbal IQ and theory of mind, which led the author to conclude the two measures would be interdependent. Using modified versions of Happé's original task, other studies later also found mindreading difficulties in higher functioning adults with autism spectrum disorders (Jolliffe & Baron-Cohen, 1999; Kaland et al., 2002). In the latter study, however, Kaland, reported a significant correlation between verbal IQ and mentalising performance among Asperger individuals. Overall, these associations between intelligence and mentalising ability raise questions about the usefulness of story comprehension tasks as tools for the assessment of social cognition.

Tasks that involve processing stimuli extracted from real life contexts are likely more ''pure.'' For example, Baron-Cohen's ''Reading the Mind in the Eyes'' test involves inferring other persons' mental states from a photograph of only their eye region (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001a). In both the original and the revised version, the test showed significant differences in mind-reading abilities between high-functioning adults on the autism spectrum and control subjects. Other investigators, using a similar task, have confirmed these findings (Kleinman, Marciano, & Ault, 2001). Studies have also demonstrated problems among able individuals on the autism spectrum in their ability to infer mental states from vocal recordings (Kleinman et al., 2001; Loveland, Tunali-Kotoski, Chen, Brelsford & Ortegon, 1995; Rutherford, Baron-Cohen, & Wheelwright, 2002).

Recently, advanced tests have been developed using video in an effort to increase test sensitivity and approximate the demands of everyday life social cognition. The ''Awkward Moment Test'' (Heavey, Phillips, Baron-Cohen, & Rutter, 2000) and the ''Empathic Accuracy Paradigm'' (Roeyers, Buysse, Ponnet, & Pichal, 2001) are most relevant to the present report. In the Awkward Moment Test, Heavey et al. (2000) showed high functioning autistic and Asperger adults and controls seven film excerpts taken from television commercials. For each film excerpt, subjects were required to answer one question on a character's mental state and one non-social question. The results revealed significant differences between the two groups in their ability to mentalise. What complicates the interpretation of these data is that the affected group also scored significantly lower on the control questions, which may reflect the difference in IQ that was observed between groups.

Roeyers et al. (2001) developed another video-based instrument, the Empathic Accuracy Paradigm. The test requires the subject to judge the feelings and thoughts of two characters that have unknowingly been filmed previously during a one-on-one conversation. The standard used for correct responses is more objective than in other tests because they are directly matched to the target character's actual subjective experience, as assessed by an interview immediately after the hidden filming. This test has proven sensitive in differentiating high-functioning individuals with pervasive developmental disorders (PDD) from control subjects. Although the test has numerous items, the mental states to be inferred are narrow in range (e.g. complex emotions or classical theory of mind concepts are missing). This is likely a result of the situation filmed: two strangers left at their own device for about 10 min waiting for a board game experiment to begin. In addition, the rate

of correct mental state answers is very low for both the affected and the control group, perhaps reflecting the difficulty in matching mental states in the relative absence of an eventful context.

The present report describes the development of another naturalistic, video-based instrument for the assessment of social cognition. The Movie for the Assessment of Social Cognition (MASC) requires study subjects to make inferences about video characters' mental states. Realization of the test entailed the development of a script, the shoot of the actual movie with actors and a professional camera team, and post processing of the film material with subsequent test-formatting. The de novo design gave us a greater level of control over the generation of mental states to be inferred. The test considers different mental state modalities (thoughts, emotions, intentions) with positive, negative, and neutral valence (Kalbe et al., in preparation). We deliberately varied the extent and quality of language, gestures, and facial expressions involved across items. We also adopted classical social cognition concepts such as false belief, faux pas, metaphor, or sarcasm to allow for a broad range of mental states to be displayed. Although the items vary in difficulty, the test was designed to be challenging so as to detect even subtle difficulties in social understanding. In this first use of the MASC, we chose to restrict the affected group studied to Asperger individuals because these individuals typically present only subtle impairments in social cognition, without any intellectual deficits. We excluded subjects that met diagnosis of high-functioning autism (HFA) because several studies seem to indicate a lesser degree of social cognitive impairments in Asperger individuals relative to HFA individuals (Jolliffe & Baron-Cohen, 1999; Dyck, Ferguson, & Shochet, 2001; Ozonoff et al., 1991). Consequently, we studied a homogeneous and able sample to assess the test's sensitivity. To minimize confounds, the control group was chosen to not significantly differ from the affected group in age, gender, education, or IQ.

In addition to the MASC, all participants received three state-of-the-art tests of social understanding as part of a broader neuropsychological battery. The inclusion of already established social cognition tests would help to further ascertain the MASC's characteristics and it's validity. We expected the MASC to clearly differentiate between diagnostic groups and, because of the pure character of the measure, we expected it to be independent of IQ.

Method

Participants

Twenty-one Asperger adults (AS; 19 men and 2 women, mean age = 41.6, SD = 10.4, range = 25–62) participated in the study. AS individuals were recruited through local support groups or were referred by specialized clinicians. Every subject underwent an extensive videotaped diagnostic interview. Based on this videotaped interview, a diagnosis of AS was made using Diagnostic Statistical Manual, 4th Edition (DSM-IV) criteria (American Psychiatric Association, 1994). Diagnostic discrepancies were resolved by consensus of one psychiatrist and two psychologists. We also utilized the Autism Diagnostic Interview—Revised (ADI-R; Lord, Rutter & Le Couteur, 1994) in 16 of the 19 Asperger subjects with available parental informants. The ADI-R is a valid and reliable semi-structured interview used for the diagnosis of autism. The instrument contains an algorithm for the diagnosis of autism as a result of probes regarding social, communication, and restricted-repetitive behavior domains corresponding to the different diagnostic criteria. For each of the three domains, a separate score is derived by summing up the items pertaining to it.

Two subjects were excluded from the study, one because the ADI-R revealed a delay in language and another failed to meet diagnostic criteria based on the taped interview. All analyses and reported results are only for those 19 individuals with a clear diagnosis of AS who have no reported language delay.

With a mean estimated Wechsler Adult Intelligence Scale (WAIS) full scale IQ of 122 (SD = 6.1, range = 111–134), and a mean of 16.7 years of education (SD = 1.7, range = 12–18) the group represented an exceptionally high-functioning sample. To assess intellectual functioning, the Shipley Institute of Living Scale (Prado & Taub, 1966) was utilized, comprising a vocabulary and an abstract thinking test. Based on a sum of the raw scores of the tests, the WAIS full scale IQ was estimated using published methods (Zachary, Paulson, & Gorsuch, 1985).

A group of 20 healthy neurotypical control subjects (18 men and 2 women, mean age = 39.9, SD = 12.6), chosen to match the patient group as closely as possible with respect to age, gender, IQ, and education, also participated in the study. Individuals in the control group were healthy volunteers participating in ongoing studies of normal aging and dementia at the NYU Center for Brain Health and had a mean IQ of 124 (SD = 6.3, range = 108–139) and a mean of 16.8 (SD = 1.4, range = 14–19) years of education.

To exclude individuals with conditions that could significantly impact on their functional ability, all subjects underwent medical (including blood work and EKG), neurologic, psychiatric, and neuroradiologic (MRI) examinations. Any present or prior evidence of significant neurologic or medical disease lead to exclusion from the study. In addition, for participants of both groups the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001b) was administered to

assess the amount of autistic traits in any one individual. The control group's mean AQ score was 15 (SD = 6, range = 6–30) and the AS group's mean score was 38 (SD = 5, range = 28–46), with the latter being well above the suggested cutoff of 32. Two individuals of the AS group scored below the cutoff. However, both individuals were clearly identified as having Asperger syndrome after administration of the ADI-R as well as the videotaped diagnostic interview.

The demographic characteristics of the participant groups are shown in Table 1. Comparisons between groups for age, education, and IQ were non-significant, whereas they were highly significant for the AQ ($p < .001$) (see Table 1).

All participants gave informed written consent and the research protocol was approved by the IRB of the New York University School of Medicine.

Measures

We administered the MASC, shortened versions of the Strange Stories Task (Happé, 1994) and the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001a), and a basic emotion recognition task (Ekman & Friesen, 1971) to all participants. Furthermore, an extensive neuropsychological test battery was given, entailing tests assessing attention, memory, and executive functions.

The MASC

The MASC was developed in collaboration with the Max Planck Institute for Neurological Research in Cologne, Germany and exists in the same format in English and in German language. The German version is currently used to assess social cognition in schizophrenia (Kalbe et al., in preparation).

The MASC requires study subjects to watch a 15 min movie about four characters getting together for a dinner party. The video is paused 46 times and questions concerning the characters' feelings, thoughts, and intentions are asked.

Theoretical Considerations

In this new test, we aimed at operationalizing social cognition through video, approximating social interactions the way they actually happen in everyday life. The storyline of the video was designed to be simple. Distracting stimuli, such as music or additional characters, were avoided. The featured characters (Sandra, Betty, Michael, and Cliff) have very different motives for partaking in an evening of cooking, dining, and playing a board game. Each develops her/his own dynamics with each of the other characters. Friendship and dating issues are the predominant themes throughout the movie. Each of the characters displays stable characteristics (traits) that are different from one another (e.g. outgoing, timid, selfish) and in the course of the evening experiences different situations that elicit emotions and mental states such as anger, affection, gratefulness, jealousy, fear, ambition, embarrassment, or disgust. The relationships between the characters were that of either strangers or friends, to vary the amount of intimacy their interactions are based on and thus represent different social reference systems on which mental state inferences have to be made. In the design of the script, varying levels of complexity were considered in creating scenes that involve the interaction of two, three, or four characters. A description of one scene with its subsequent question is given in the Appendix.

Not only did the de novo design enable us to adopt traditional social cognition concepts such as first and second order false belief, deception, faux pas, persuasion,

Table 1 Mean scores (M), medians (MD), standard deviations (SD), and ranges of the demographic characteristics of both groups

	Shipley					
	Age	Education	Vocabulary	Abstract thinking	WAIS IQ	AQ
Asperger (n = 19)						
M	**41.6**	**16.7**	**35.3**	**35.1**	**122**	**37.8**
MD	41	18	37	34	121	39
SD	10.4	1.7	3.7	3.1	6	4.9
Range	25–62	12–18	28–40	28–40	111–134	28–46
Control (n = 20)						
M	**39.9**	**16.8**	**36.5**	**35.7**	**124**	**15.4**
MD	42.5	17	37.5	35	124	14.5
SD	12.6	1.4	2.5	4.1	6	6.4
Range	22–60	14–19	32–40	24–40	108–139	6–30
p Value	.63 [a]	.98[b]	.37[b]	.27[b]	.34[a]	< .001[b]

[a] *t* Test

[b] Mann–Whitney *U* test

metaphor, sarcasm, or irony, it also gave us the opportunity to implement for the first time a more multidimensional approach. It allowed us to ensure proper coverage of sub-divisions of different mental state modalities, valence, and degrees of language involvement in the social cognition process.

The different mental state modalities that were taken into consideration are ''emotions,'' ''thoughts,'' and ''intentions'' (Kalbe, Brand, Fleck, & Kessler, 2002). Questions were implemented in the format of: ''What is Sandra feeling?'' ''What is Michael thinking?'' or ''What is Betty's intention?'' Of these mental state modalities, the category ''emotion'' comprises 17 items, the category ''thought'' 7 items, and the category ''intention'' 18 items.

To allow the assessment of emotional mental state items of different valence (positive, negative, neutral), we designed scenes in which the characters express negative feelings like disgust, anger, or fear, as well as positive emotions like joy or affection. Of the 17 emotional mental state items, 13 are of negative valence, 2 of positive valence, and 2 are neutral. These frequencies approximate distributions previously considered by others (e.g. Ekman, 1999).

In addition to the above mentioned mental state modalities, some of the items varied as to their conversational content. Specifically, items were designed to be verbal (19 items) or non-verbal (16 items), with the verbal items to be taken literally (10 items) or not literally (those containing figurative speech and other aspects of pragmatics, 9 items). The non-verbal category provided items to assess the recognition of facial expressions (6 items), as well as a broader category that requires the interpretation of body language and gestures (10 items). Single items may cover more than one domain, e.g. item 5 pertains to both the mental state category ''intentions'' and the conversational category ''figurative speech.''

A more detailed discussion on the multidimensional features of the test is beyond the scope of this paper and will be reported on elsewhere.

Development and Realization

The storyline of the MASC is based on a script developed by the first two authors. The writing process followed guidelines provided by Field, Meyer, and Witte (2001) on the development of screenplays. Among other steps, it involved designing characters, the plot, and a final dialogue form. For the purpose of creating a mindreading tool, it was important to develop ''whole'' characters. This required that, for each character, a careful detailing of their ''worlds,'' including background information such as profession, lifestyle, hobbies, family history, self-assessment, personality traits, and needs be created. We assigned basic needs that do not vary over the course of the video (e.g. Michael is romantically interested in Sandra, Betty wants to be a good friend to Sandra), as well as sub-needs that vary within a character in different situations (e.g. being polite, getting back at somebody). These needs represent a foundation for the characters' mental states which have to be appreciated by the study subjects. Through the specific implementation of ambiguity between basic and sub-needs (e.g. Michael gets back at Sandra although he wants to date her), we created items that pose particularly challenging demands on social cognitive functioning.

We attached great importance to a high quality technical production in shooting the movie. A trained team composed of a professional cameraman, sound engineer, and four actors belonging to a long standing theatre group accompanied by their director, helped accomplish this goal within a three-day shoot. The raw material of the film was captured from Digital Video (DV) to a digital format on a computer and cutting, light-, and sound post-processing was done by a film editor using the software ''Final Cut'' and ''Quick Time Pro.'' The video was then saved as an MPEG file format and was cut into 46 segments that represent the individual items of the test. The segments and their subsequent questions, as well as the instructions, were inserted into a ''PowerPoint'' presentation that can be shown on a regular PC or notebook.

The original version of the movie was produced in Dortmund, Germany with German speaking actors. For the English version the script was translated into English by a language professional and the 15 min film material that went into the test was dubbed by a team of sound and theatre professionals in a sound studio. In several sessions with each character being voiced-over separately, voices and all background sounds were recorded and processed in ''ProTool.'' After merging the picture file with the new sound track, the film was formatted in the same way as the German version. Study participants are informed about the dubbing before the test administration. However, post experimental questioning revealed that the dubbing was in no case reported as interfering and in most cases reported as not remarked upon.

Administration

The MASC is administered by a tester who controls the presentation of the test's slides. The testing starts with a slide instructing the subjects that they are going to watch a 15 min film and that they should try to understand what the characters are feeling and thinking. The tester navigates through several slides containing instructions and then through the entire test using the mouse or space bar. As part of the first slides, the four characters are introduced in the form of photographs and names. After that, participants are instructed that the film shows these four people getting

together for a Saturday evening and that the movie will be stopped at various points and questions will be asked. Subjects are told to try to imagine what the characters are thinking or feeling at the very moment the film is stopped. Following the instructions, the 46 video segments are presented, each followed by a question in the same format (e.g. ''What is Betty feeling/thinking/intending to do?''). Subjects are instructed to respond verbally to each question. We chose an open-answer format for this first study to minimize correct answers through guessing and to collect qualitative data that will be analyzed and reported on elsewhere (Fleck et al., in preparation). Administration of the MASC takes approximately 45 min.

We recently completed a multiple-choice format for the test (Fleck et al., in preparation). The incorrect answers that were given in the present study were used as models to construct distractor answers. Specifically, the multiple-choice format allows differentiation of three different types of mistakes that reflect (1) mental state inferences that are ''insufficient'' and (2) ''too excessive''. In addition, one type of distractor answer reflects (3) non-mental state inferences (i.e., physical causation). This more user-friendly version of the MASC is also available to other researchers, by request to the corresponding author.

Scoring

All responses are recorded on audio tape and scored later, using a standardized scoring key that provides several examples for correct and incorrect answers for each item. The scoring key was derived following a two-step procedure. First, the final version of the video was judged on agreement between how mental states were intended in the script and how the actual mental states were depicted by the characters. Only if there was agreement was the item included in the final test format. As a next step, preliminary data was collected from 30 healthy control subjects and items were tested for feasibility. None of the 46 items initially chosen had to be rejected.

Correct responses are scored as one point, and incorrect responses as zero points. An overall score as well as the different sub-component scores are derived. In the present study, all ratings were done by the first author. To ascertain the reliability of the scoring, a second rater independently scored the video-test for five patients and five control subjects. The consistency (ICCs) for the 10 ratings was .99 (.98 for the AS group and .94 for the control group), indicating high interrater reliability. Examples of scoring criteria for one item are given in the Appendix.

To control for memory and general comprehension effects, four control questions are asked following the video-test presentation (e.g. which beverages did the characters drink?). These questions are scored as one, .5, or zero points.

Strange Stories Task

The test material was comprised of eight theory of mind stories and two control stories of Happé's (1994) original Strange Stories Task. Subjects were asked to read, on separate pages, short passages of text and to answer a question that was presented on a subsequent page for each passage. For the theory of mind questions, subjects were required to infer a character's mental state whereas the control questions were asked for the interpretation of physical events. The theory of mind stories concerned two examples of double bluff, persuasion, irony, and white lie. Scoring was done using a detailed rating scheme that was communicated personally to the authors by Happé and which was in parts outlined previously (Happé, Winner & Brownell, 1998). Answers were scored 2 if they were fully and explicitly correct, 1 if they were partially or implicitly correct and 0 if they were incorrect. Because judging whether an answer is correct or not involves subjectivity, a second rater independently scored the stories for five patients and five control subjects. The consistency (ICCs) for the 10 ratings was .99 (1.0 for the AS group and .93 for the control group), indicating a high degree of concordance.

Reading the Mind in the Eyes Test

The revised version of the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001a) involves inferring the mental state of a person just from the information conveyed in photographs of that person's eyes. Baron-Cohen's task was shortened to include 24 of the original 40 items. Participants were asked to pick for each pair of eyes one out of four mental state descriptors (e.g. interested, hostile), where these descriptors varied with each item. They were also instructed to indicate the gender of the person in the picture to control for deficits in general face or social perception. The test was scored by adding up the number of mental state and gender attributions correctly identified.

Basic Emotion Recognition

To test the ability of judging emotions expressed in faces, participants were given a series of 28 pictures of facial affect by Ekman and Friesen (1971). With each face, a wordlist of six basic emotional states (happiness, sadness, fear, disgust, anger, surprise), intermixed with the word ''neutral'' was displayed in random order and the subject was required to choose the one word that they considered to best describe what the person in the photo was feeling. The criterion measure was again ascertained by totaling the number of items correctly identified.

We chose to add this facial emotion recognition test to the battery of social cognition measures because emotions

are considered a more structured subset of mental states and the recognition of emotions from facial cues is a key mindreading element.

Executive Functions, Attention, Visual Processing, and Memory

To control for possible confounds from other areas of neuropsychological functioning, a variety of measures were applied. Specifically, to test executive functions, the Stroop Test (Stroop, 1935), a verbal fluency test (Horn, 1962), and the Trail Making Test (Reitan & Wolfson, 1993) were administered. The Digit Symbol Substitution Test (Wechsler, 1955) was used to assess attention and the digit span forward and backward (Wechsler, 1987) to assess short-term and working memory. In addition, the subtests Logical Memory and Visual Reproduction from the Wechsler Memory Scale—Revised (Wechsler, 1987) were used to test verbal and visual declarative memory. To control for visual processing deficits, a mental rotation test and a spatial visualization test were given (Horn, 1962).

Procedure

All participants were tested individually at the Center for Brain Health, NYU School of Medicine in a quiet room by trained examiners. Because participants of the control group were recruited from different ongoing studies, part of their cognitive testing was done in conjunction with these research agendas. However, all social cognition tests and most of the other neuropsychological measures were administered as part of a separate cognitive session and were given in the same order the AS group received them in.

Statistical Analysis

The data were analyzed using the Statistical Program for Social Sciences version 11.0 (SPSS, Chicago, Ill) and MedCalc version 7.2 (MedCalc, Mariakerke, Belgium). All variables were tested for normal distribution with the Kolmogorov–Smirnov test. Where appropriate, independent *t* tests were used to test for between-group differences. Because the data for most measures were not normally distributed, we conducted non parametric Mann–Whitney *U* tests to assess group differences.

We applied Receiver Operating Characteristic (ROC) curves for the social cognition measures to test for classification accuracy. ROC curves entail plotting the balance between the sensitivity and specificity of a test while systematically moving the cut score across its full range of values. In a ROC curve plot, the diagonal line demonstrates the ''random ROC,'' which reflects a test with zero discriminating power. The accuracy of the ROC curve is quantified by calculating the area under the curve (AUC). An AUC of .50 indicates that a test's diagnostic performance is equal to chance, whereas an AUC of 1.0 indicates perfect diagnostic performance. In addition, Pearson correlations were used to assess associations between the administered measures.

Results

MASC

The difference in correct mental state inferences for the MASC was highly significant (Mann–Whitney: $U = 9.5$, $p < .001$), indicating greater difficulties in the Asperger group. There was no difference, however, between the two groups for the control questions (Mann–Whitney: $U = 189$, $p = .94$).The results on the MASC and the other social cognition tests are shown in Table 2.

Reading the Mind in the Eyes Test

There was a highly significant difference between the groups in the number of mental states correctly ascribed (Mann–Whitney: $U = 58$, $p < .001$), with the control group showing better performance than the Asperger group. In line with Baron-Cohen (Baron-Cohen et al., 1997; Baron-Cohen et al., 2001a), we did not find a difference for the groups in the gender control task (Mann–Whitney: $U = 189$, $p = .97$).

Basic Emotion Recognition

The *t* test examining between-group differences on the Emotion Recognition task yielded a significant *t* value ($t = 3.2$, $p < .01$). The analysis showed that the Asperger individuals performed more poorly than the control group in identifying basic emotions from photographs.

Strange Stories Task

The Mann–Whitney *U* testing for between-group differences on the Strange Stories Task yielded significant differences for the number of correct mental justifications (Mann–Whitney: $U = 126$, $p < .05$), but not for the number of correct physical justifications (Mann–Whitney: $U = 179$, p .77).

Executive Functions, Attention, Visual Processing, and Memory

None of the above listed tests yielded significant between-group effects.

Table 2 Performance of the AS group and control group on the MASC, the Reading the Mind in the Eyes Test, basic emotion recognition, and the Strange Stories Task

Max. scores:	MASC		Reading the Mind in the Eyes		Emotion Recognition	Strange Stories Task	
	Test questions 46	Control questions 4	Eyes task 24	Gender task 24	Identified 28	Mental stories 16	Physical stories 4
Asperger (n = 19)							
M	**24.4**	**3.9**	**16.1**	**22.5**	**22.6**	**14.2**	**3.8**
MD	26	4	15	23	23	15	4
SD	5.9	.2	3.1	.8	2.6	2.7	.5
Range	13–33	3–4	10–21	21–24	16–27	7–16	2–4
Control (n = 20)							
M	**34.8**	**3.9**	**20.0**	**22.4**	**25.3**[c]	**15.7**	**3.9**
MD	35	4	20	23	26	16	4
SD	2.7	.1	1.8	1.2	2.1	.5	.2
Range	30–39	3.5–4	15–23	19–24	21–28	15–16	3–4
p Value	<.001[b]	.94[b]	<.001[b]	.97[b]	<.01[a]	.04[b]	.77[b]

[a] *t* test

[b] Mann–Whitney *U* test

[c] Six controls were missing this test, due to its later inclusion in the test battery

Receiver Operating Characteristic (ROC) Curves

ROC curves determined the relative value of the four social cognition measures in their ability to make a diagnostic group distinction. The area under the ROC curves for the social cognition measures were .98 for the MASC, .86 for the Reading the Mind in the Eyes Test, .79 for the emotion recognition test, and .65 for the Strange Stories Task (see Fig. 1).

Pairwise comparisons of the areas under the ROC curves identified the MASC as significantly more accurate in discriminating the AS group from the control group than the Reading the Mind in the Eyes Test ($\Delta = .13$, $p < .05$), the emotion recognition test ($\Delta = .19$, $p < .05$), and the Strange Stories Task ($\Delta = .31$, $p < .01$).

Correlations

To assess relationships between intelligence and mentalising abilities, correlation analyses were performed (see Table 3). Among the individuals in the AS group there was a significant correlation between the vocabulary test and the Strange Stories Task ($r = .48$, $p < .05$). In the control group, we observed associations between the abstract thinking test and the basic emotion recognition task ($r = .63$, $p < .05$). There were no significant associations between the MASC and any of the IQ measures for either group.

Intercorrelations of the social cognition measures revealed a significant association between the MASC and the Strange Stories Task ($r = .47$, $p < .05$) in the AS group, and between the MASC and the basic emotion recognition task in the control group ($r = .72$, $p < .01$). It is interesting to note that these were the only significant associations between the social cognition measures, possibly indicating that the MASC shares features with these two mentalising tests, whereas this is not the case between the other three social cognition measures. To further validate the MASC, we performed correlation analyses between all the social cognition tests and the score of the ADI-R social

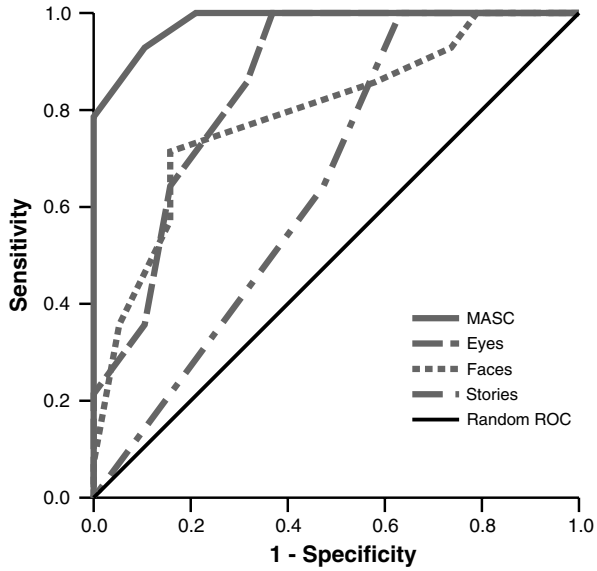

Fig. 1 ROC curves for all social cognition measures (MASC, Eyes = Reading the Mind in the Eyes Test, Faces = basic emotion recognition test, Stories = Strange Stories Task) and a reference curve depicting a system making random predictions (Random ROC). Areas under the curve were, in the order listed, 0.98, 0.86, 0.79, and 0.65, respectively

Table 3 Intercorrelations between social cognition tests (Eyes = Reading the Mind in the Eyes Test, Faces = basic emotion recognition test, Stories = Strange Stories Task) and correlations with Shipley IQ measures

	Age	IQ measures			ADI-R[b]	Social cognition measures		
		Vocabulary	Abstract thinking	Estimated WAIS IQ	Social domain	Eyes	Faces	Stories
Asperger (n = 19)								
MASC	.159	.183	.201	.117	−.533***	−.273	−.078	.471*
Eyes	.311	.374	−.150	.278	−.080		.375	.043
Faces	.401	.405	−.353	.255	−.197			.036
Stories	.052	.479*	.224	.193	−.285			
Control (n = 20)								
MASC	−.198	.285	.251	.006		.078	.715**	.272
Eyes	−.080	.059	−.245	−.173			−.353	.189
Faces[a]	−.175	.331	.631*	.470				.328
Stories	.214	.089	.223	.216				

[a]Six controls were missing this test, due to its later inclusion in the test battery

[b]Four Asperger individuals were missing the ADI-R, due to unavailable parental informants

*$p < .05$, **$p < .01$, ***$p < .1$

domain for the Asperger individuals. We observed a trend towards statistical significance for the relationship between the MASC and the ADI-R ($r = -.53$, $p < .1$). The other social cognition tests had much lower levels of associations. All relationships are listed in Table 3.

We also ran correlations between the MASC and the other neuropsychological tests administered and did not find any significant associations in the AS group for either one of the executive functions, attention, visual processing, or memory variables. However, for the control group, we observed associations between the MASC and both delayed paragraph recall ($p < .05$) and the spatial visualization test ($p < .05$).

MASC Internal Consistency

Internal consistency was assessed by calculation of Cronbach's alpha, which revealed highly satisfactory values. A value of alpha = 0.70 or above is considered to be acceptable (Nunnally, 1978). Alpha was 0.84 for the total scale, with the range in internal consistency, as measured by alpha if item deleted, being 0.82 to 0.84. This alpha if item deleted statistic indicated that removal of any item would result in a lower alpha for the MASC, thus indicating the utility of all items.

MASC Test–Retest Reliability

Five Asperger individuals and 5 control subjects were asked to come back for a second administration of the MASC to establish the test–retest reliability. The second administration took place one to 12 months after the first one, with an average interval of 4.6 months for the AS group and 3.6 months for the control group. The control group's mean scores for the first and second administration were 34 (SD = 2.3) and 35 (SD = 2.7), respectively and the AS group's mean scores were 21.4 (SD = 6.5) and 24 (SD = 5.1), respectively. We computed intraclass correlation coefficients (ICCs) and observed high agreement between administrations for the whole sample (ICC = .97) and also for the groups individually (AS: ICC = .92 and NC: ICC = .89).

Discussion

In the present study we introduced the MASC, a new tool for the assessment of mindreading abilities in individuals with a diagnosis of Asperger syndrome. Asperger and control participants were administered the MASC and three additional tests of social cognitive functioning as part of a neuropsychological test battery. Overall the results confirmed a selective impairment of social inferring in the affected group. IQ, executive function, memory, attention, and visual processing were not different between the study groups.

The MASC proved to be sensitive in detecting mindreading difficulties in the AS group. The test required subjects to attribute mental states to movie characters in an everyday life relevant context. All mentalising tests separated the two groups. However, comparisons of the areas under the ROC curves demonstrated that the MASC was the superior test in discriminating the AS group from the control group. Furthermore, the MASC was the only test that showed a trend towards statistical significance for a negative association with the social domain of the ADI-R, indicating that the more severely an individual is affected, the poorer is the performance on the MASC. Our de novo design of the MASC allows for a very broad range of mental states tested,

including traditional theory of mind concepts. The observation of convergent results with two other social cognition tests, as well as the associations between the ADI-R, provides evidence for the validity of the new video-test.

In line with Baron-Cohen et al.'s findings (2001a), our Asperger subjects performed less well than control subjects on the Reading the Mind in the Eyes Test. In addition, we also found that our AS group was impaired on the basic emotion recognition task. There has been a lack of consensus as to whether AS individuals have deficits in the recognition of basic emotions from facial stimuli; with negative (Baron-Cohen, Wheelwright, & Jolliffe, 1997; Prior, Dahlstrom, & Squires, 1990) and positive reports (Macdonald et al., 1989; Njiokiktjien et al., 2001; Scott, 1985). In an earlier study that, like the current study, used both the Reading the Mind in the Eyes Test and a simple emotion recognition task, Baron-Cohen et al. (1997) found HFA and AS individuals impaired on the first, but not the latter test. However, in that study he utilized an easier emotion recognition task (two choices) rather than the seven choices (six basic emotional and a neutral mental state) of the present study. In fact, Baron-Cohen's task was so easy that members of both groups scored ceiling.

Given the importance of understanding facial expressions in social interactions, and considering the numerous reports describing atypical face processing in autism spectrum disorders (e.g. Hobson, Ouston, & Lee, 1988; Pelphrey et al., 2002; Teunisse & de Gelder, 2003), it would be important to ascertain the exact nature of the deficits. One of the more critical questions to answer will be whether autistic conditions involve deficits in the visuospatial aspects of face perception (i.e. less expressive areas of face are centers of attention such as mouth rather than the eyes) or rather deficits in the interpretation of expression (i.e. lack of mental state concepts) (see also Grossman, Klin, Carter, & Volkmar, 2000). To clarify this issue future research should systematically control all three variable dimensions: complexity of mental states (basic emotions versus complex mental states), part of face (e.g. eye region versus mouth region), and answer format (e.g. two-choice versus multiple-choice format). Although this study was not designed to address this issue directly, we found hints of a basic impairment in the interpretation of facial expressions. We found no significant differences in the area under the ROC curves for the emotion recognition task and the Reading the Mind in the Eyes Test and one possible implication of this result might be that more basic difficulties with the recognition of simple emotions underlie the problems seen in the more complex mental state inferences required in the Reading the Mind in the Eyes Test.

Our finding that AS individuals are impaired on the Strange Stories Task is consistent with Happé's (1994) observations in able autistic adults. However, in the present study the between-group difference was less pronounced than for the other social cognition measures. In addition, the Strange Stories Task was the only test of social understanding that showed associations with verbal IQ, corroborating findings from Kaland et al. (2002) and identifying story comprehension as a less pure format for assessing social cognition.

We found no differences between our groups for neuropsychological measures of executive functions, attention, memory, visual processing, or IQ. Furthermore we found no associations between these neuropsychological measures and MASC scores among the AS group. These results are of considerable importance given discussions about the ''pureness'' of video formats in the assessment of social cognition. Although video formats such as the MASC approximate everyday social interactions more adequately than static pictures or story formats, it has been argued legitimately that they also involve executive functions and central coherence and hence are not ''pure'' social cognition tasks (Baron-Cohen et al., 1997; Heavey et al., 2000; Roeyers et al., 2001). In research involving individuals on the autism spectrum, even more attention needs to be brought to this argument because central coherence and executive functions have been reported as impaired in affected individuals (Frith & Happé, 1994; Hughes, Russell, & Robbins, 1994). When designing the MASC, we tried to minimize demands on executive functions and central coherence. We accomplished this by avoiding distracting or prompting stimuli such as music, direct camerawork, a complete or targeted storyline, and fast changing scenes that existing TV clips so often present with. We also implemented control questions to consider confounds. However, these questions proved too easy and resulted in ceiling performance by both groups. We agree with Heavey et al. (2000) that control questions, in order to control for executive functions and central coherence, should be formulated in a way so as to be inferential and involve the recollection and integration of film information without demanding social understanding required by the test questions. We plan to implement more challenging control questions in the near future. However, at this point we have to rely on results from the additional neuropsychological measures administered that seem to indicate that performance on the MASC is largely independent of other cognitive areas.

We found the MASC to be a reliable instrument. Not only did the test prove to have high interrater reliability and internal consistency, the results also seem to be highly stable over time. Given that none of the subjects that were asked to come back to take the MASC again underwent any kind of social cognitive intervention since their first visit the consistency over time qualifies the MASC as a potential tool to monitor treatment efforts.

Intercorrelations of the social cognition measures revealed that there were no associations between the Strange Stories Task, the Reading the Mind in the Eyes Test and the basic emotion recognition test, suggesting that the tests may assess different aspects of social cognition. The important implication from these results is that social cognition is a multifaceted construct.

We found that performance on the MASC was associated with performance on the Strange Stories Task in the AS group and with performance in the basic emotion recognition test in the control group. This seems to support the idea that the MASC may, in part, measure similar functions as the two other mindreading tests. These results are not unexpected given that the MASC was designed to cover some of the conceptual areas covered by those tests. Namely, the MASC was constructed so as to incorporate classical social cognition concepts and pragmatic language items such as sarcasm, false belief, or deception that are also covered by the Strange Stories Task. Similarly, the MASC contains items that require the subject to read, relatively independent of context, facial expressions in order to correctly identify a character's mental state, thus mirroring the basic emotion recognition task. However, it is interesting to note that the relationships between both the MASC and the Strange Stories Task and the MASC and the emotion recognition test are different for the AS and the control group. Although speculative at this point, it is possible that while unaffected individuals rely heavily on facial cues in appreciating others' mental states (which explains the association between the facial emotion recognition test and the MASC in the control group), a core impairment in reading facial expressions accompanying AS requires individuals to use compensatory strategies such as interpreting verbally communicated information (which explains the association between the Strange Stories Task and the MASC in the AS group). Findings from a study by Grossman et al. (2000) can, in part, be seen as a corroboration of this assumption. The authors found a group of AS children to be impaired in their ability to recognize simple facial emotions only when faces were paired with mismatching emotional words. There were no differences from a control group when faces were paired with matching or irrelevant words. The authors interpreted the results as indicator for a bias towards visual-verbal over visual-affective information in the AS individuals (words over faces). Also in line with this argument are results from a study using the Empathic Accuracy Paradigm (Roeyers et al., 2001), a video-based instrument similar to the MASC. In their research, the investigators used this test along with a stories task and an eye photographs task in a group of adults with PDD. Intercorrelations of these tests showed a significant association between the eye photographs task and the Empathic Accuracy Paradigm in the control group, and a trend between the stories task and the Empathic Accuracy Paradigm in the PDD group, with reverse associations not having occurred.

In the present study, we observed differing associations between abstract thinking and facial emotion recognition for the two groups. In the control group, the abstract thinking subtest of the Shipley Institute of Living Scale was found to be positively related to emotional face recognition. This may indicate that neurotypical individuals use similar cognitive capacities for face ''decoding'' as for abstract concept formation, which has the specific component skills of cognitive flexibility, attention to detail, and analysis and synthesis (Zachary et al., 1985). Although the AS individuals were as proficient on abstract concept formation as the control group, they do not seem to apply these cognitive skills to face stimuli. This may, again, reflect differences in processing strategies in AS individuals.

Another correlation with opposite patterns for the two groups was observed for the Reading the Mind in the Eyes test and the emotional face recognition. Although the individual results were not significant, the magnitude of the difference between the groups (Control: $r = .-353$, Asperger: $r = .375$) warrants a cautious interpretation. It seems possible that for neurotypical controls the different stimuli formats of the two tasks (whole face versus eye region only) caused the varying performance. While trying to decode a whole face with abstract and analytical thinking abilities (the aforementioned association between abstract thinking and face recognition), the information provided by only the eye region might simply be too limited to use such strategy (the abstract thinking subtest was not found to be related to the Reading the Mind in the Eyes test). In that case, control subjects might rely on intuition or ''gut-feeling'' to perform the task. Thus, a negative association between tasks could be reflective of these different processing strategies (abstract thinking versus intuition). In contrast to controls, who are expert face readers, for AS individuals task format might not be of relevance. Their innate difficulty in processing faces and facial features might be an all-determining factor for task performance. This may prevent more subtle effects of task format from showing an effect in the AS group.

We believe that the MASC is an important addition to the available instruments to assess social cognition. With that being said, there are some issues about the MASC that deserve further discussion. First, all four featured characters of the MASC are of roughly the same age: in their mid-thirties. Age serves here as a context in which mental state attributions are to be made. Since the appropriateness of certain behaviors has changed over the last decades (e.g. men offering help with the cooking) and use of language may vary with different generations, it is likely that the social interactions depicted in the test are more easily

understood by individuals of the same age group. For the groups reported here age is not likely to be a concern since the characters' age mirrors closely the mean age of our subjects. In addition, we did not find any associations between age and social cognitive performance in the MASC. However, future video tests should consider involving characters of a wider age range.

Another possible issue of concern for the MASC is that it is relatively time consuming. Due to the open answer format, a test session requires 45 min for completion and the subsequent scoring requires a trained rater who is familiar with the scoring procedure and scoring key. However, the recently completed multiple-choice format of the test (Fleck et al., in preparation) has shorter administration times (approximately 30 min) and automated scoring.

There are a number of other possible research avenues that the MASC will help explore. Differences in response patterns within individuals with AS and cross-cultural effects on social cognitive functioning are only two of them. In future work, we plan to add a direct brain assessment utilizing MRI. By relating scores of subcomponents obtained with the MASC to volumes of brain structures thought to be involved in social cognition, we will be in a good position to contribute towards identifying the brain underpinnings of social cognitive impairments.

Conclusion

Very able AS individuals were found to have selective impairments in social cognition. Out of four measures of social understanding used, the newly developed video-test the MASC had the greatest sensitivity in differentiating AS individuals from control subjects. Intercorrelations of the social cognition measures point to a possible multidimensionality of the cognitive construct and also suggested atypical social processing strategies for AS individuals. Future research is needed to address these questions in more detail.

Acknowledgments This research was funded by a grant from the National Alliance for Autism Research (NAAR). It was completed partially toward the first author's Ph.D. dissertation at the University Bielefeld, which was supported with a training grant by the Cusanuswerk, Germany. The development of the MASC was, in part, supported by the Max-Planck-Institute for Neurological Research and Köln Fortune, Cologne, Germany. We are grateful to the participants and their families for volunteering for the study and we thank the great number of people who donated their time and dedication to the project. Our special thanks goes to Jonathan Bepler.

Appendix:

Example scene and scoring criteria:

Scene 20:

Picture 1 Cliff is the first one to arrive at Sandra's house for the dinner party. He and Sandra seem to enjoy themselves when Cliff is telling about his vacation in Sweden (Printed with permission)

Picture 2 When Michael arrives, he dominates the conversation, directing his speech to Sandra alone (Printed with permission)

Picture 3 Slightly annoyed by Michael's bragging story, Sandra shortly looks in Cliff's direction and then asks Michael: ''Tell me, have you ever been to Sweden?'' (Printed with permission)

Question: Why is Sandra asking this?

Examples for correct answers: To change to the topic that Cliff talked about before so that he gets involved again; to redirect the conversation to Cliff; to integrate Cliff; to reconnect with Cliff.

Examples for incorrect answers: To hear if Michael also has something interesting to say about Sweden; to see which of the two guys has a cooler story to tell; to see if Michael can corroborate Cliff's story; she liked the Sweden topic better than the current one; to compare the two; to loosen Michael up, the Sweden topic also worked for Cliff.

References

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

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

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

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, Massachusetts: MIT Press.

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

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

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 29*, 407–418.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001a). 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, and Allied Disciplines, 42*, 241–251.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a ''language of the eyes''? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*, 311–331.

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

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

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience, 1*, 27–51.

Dyck, M. J., Ferguson, K., & Shochet, I. M. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests? *European Child & Adolescent Psychiatry, 10*, 105–116.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion*. Sussex, UK: John Wiley & Sons.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Personality, 17*, 124–129.

Field, S., Meyer, A., & Witte, G. (2001). *Drehbuchschreiben für Fersehen und Film*. München: Ullstein Verlag.

Fleck, S., Dziobek, I., Rogers, K., Kalbe, E., Kessler, J., Mielke, R., Daum, I., Wolf, O. T., & Convit, A. (in preparation). MASC-MC. The Movie for the Assessment of Social Cognition in a new multiple choice answer format: A study with adults with Asperger syndrome.

Frith, U., & Happé, F. (1994). Autism: Beyond ''theory of mind''. *Cognition, 50*, 115–132.

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

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*, 101–119.

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

Happé, F., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology, 34*, 358–362.

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

Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *The British Journal Of Psychology, 79*(Pt 4), 441–453.

Horn, W. (1962). *Leistungsprüfsystem: LPS*. Göttingen: Hogrefe.

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 33*, 3–13.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia, 32*, 477–492.

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., Moller-Nielsen, A., Callesen, K., Mortensen, E. L., Gottlieb, D., & Smith, L. (2002). A new 'advanced' test of theory of mind: Evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 43*, 517–528.

Kalbe, E., Fleck, S., Brand, M., Dziobek, I., Ruhrmann, S., & Kessler, J. (in preparation). ''Too much ToM'' in schizophrenic patients? Or: The necessity to differentiate the concept of Theory of Mind (ToM) and its disturbances.

Kalbe, E., Brand, M., Fleck, S., & Kessler, J. (2002). A differentiation of the construct of Theory of Mind: Theoretical considerations, corresponding tests, and preliminary data [Zur Fraktionierung des ToM-Konstrukts: theoretische Ueberlegungen, Testverfahren und erste Ergebnisse]. In M. Baumann, A. Keinath & J. F. Krems (Eds.), *Experimentelle Psychologie. Abstracts der 44. Tagung experimentell arbeitender Psychologen*. Regensburg: S. Roderer Verlag.

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., & Volkmar, F. R. (1997). Asperger Syndrome. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 94–122). New York: John Wiley & Sons.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pevasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

Loveland, K., Tunali-Kotoski, B., Chen, R., Brelsford, K., & Ortegon, J. (1995). Intermodal perception of affect by persons with autism or Down syndrome. *Development and Psychopathology, 7*, 409–418.

Macdonald, H., Rutter, M., Howlin, P., Rios, P., Le Conteur, A., Evered, C., & Folstein, S. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 30*, 865–877.

Njiokiktjien, C., Verschoor, A., de Sonneville, L., Huyser, C., Op het Veld, V., & Toorenaar, N. (2001). Disordered recognition of facial identity and emotions in three Asperger type autists. *European Child & Adolescent Psychiatry, 10*, 79–90.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Ozonoff, S., Rogers, S. J., & Pennington, B. F. (1991). Asperger's syndrome: Evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 32*, 1107–1122.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*, 249–261.

Pilowsky, T., Yirmiya, N., Arbelle, S., & Mozes, T. (2000). Theory of mind abilities of children with schizophrenia, children with autism, and normally developing children. *Schizophrenia Research, 42*, 145–155.

Prado, W. M., & Taub, D. V. (1966). Accurate predication of individual intellectual functioning by the Shipley-Hartford. *Journal of Clinical Psychology, 22*, 294–296.

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

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders, 20*, 555–567.

Reitan, R. M., & Wolfson, D. (1993). The Halstead Reitan Neuropsychological Test Battery: Theory and clinical interpretation. Tuscon, AZ: Neuropsychology Press.

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, and Allied Disciplines, 42*, 271–278.

Rutherford, 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.

Scott, D. W. (1985). Asperger's syndrome and non-verbal communication: A pilot study. *Psychological Medicine, 15*, 683–687.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 643–662.

Teunisse, J. P., & de Gelder, B. (2003). Face processing in adolescents with autistic disorder: The inversion and composite effects. *Brain and Cognition, 52*, 285–294.

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

Wechsler, D. (1955). *Wechsler adult intelligence scale*. New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler memory scale-revised*. San Antonio: Psychological Corporation/Harcourt Brace Javanovich.

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

Wing, L. (1997). Syndromes of autism and atypical development. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 148–170). New York: John Wiley.

Zachary, R. A., Paulson, M. J., & Gorsuch, R. L. (1985). Estimating WAIS IQ from the Shipley Institute of Living Scale using continuously adjusted age norms. *Journal of Clinical Psychology, 41*, 820–831.