ORIGINAL PAPER

# Theory of Mind in Adults with HFA and Asperger Syndrome

Annelies A. Spek · Evert M. Scholte ·
Ina A. Van Berckelaer-Onnes

**Abstract** Theory of mind was assessed in 32 adults with HFA, 29 adults with Asperger syndrome and 32 neurotypical adults. The HFA and Asperger syndrome groups were impaired in performance of the Strange stories test and the Faux-pas test and reported more theory of mind problems than the neurotypical adults. The three groups did not differ in performance of the Eyes test. Furthermore, correlations between the Eyes test and the three other theory of mind tests were low or absent. Therefore one can question the ability of the Eyes test to measure theory of mind. Of all theory of mind tests used, the self-report questionnaire had the largest discriminating power in differentiating the two disorder groups from the neurotypical group.

**Keywords** HFA · Asperger syndrome · Theory of mind · EQ · Faux-pas · Strange stories

## Introduction

In autism research there are three leading cognitive theories which describe impairments in 'theory of mind' (Baron-Cohen et al. 1985), 'central coherence' (Frith 1989, 2003) and 'executive functioning' (Ozonoff et al. 2005; Rumsey 1985). This paper examines theory of mind and is part of an ongoing study to assess the three above-mentioned

A. A. Spek (✉)
GGZ Eindhoven, Boschdijk 771, P.O. Box 1418,
5626 AB Eindhoven, The Netherlands
e-mail: aa.spek@ggze.nl

E. M. Scholte · I. A. Van Berckelaer-Onnes
Department of Clinical Child and Adolescent Studies,
Leiden University, Leiden, The Netherlands

cognitive domains in adults with HFA or Asperger syndrome. Although theory of mind has been studied extensively in children with autism (Baron-Cohen et al. 1985; Baron-Cohen 2000; Frith 2003), studies that examined theory of mind functioning of adults with HFA and Asperger syndrome are limited. Furthermore, previous studies in adults with ASD used both neuropsychological tests and self-reports to assess deficits in theory of mind, although the relationship between these two measurement methods was never investigated. To fill this gap, in the present study theory of mind is assessed in adults with HFA or Asperger syndrome using both neuropsychological tests and self-reports.

Theory of Mind in Autism

The term 'theory of mind' was introduced in psychology by Premack and Woodruff (1978) to describe the ability of a person to attribute mental states to oneself and others and to predict the behavior of others based on their mental states. Research throughout the years has shown that children, and to some extent also adults, with ASD experience problems in theory of mind (Baron-Cohen 2000; Bowler 1992; Frith 1989; Happé 1994; Kaland et al. 2002; Leslie 1987; Ozonoff et al. 1991b; Ponnet et al. 2004). In theory of mind functioning, a distinction is made between the different levels of theory of mind (Baron-Cohen 2000). First order theory of mind involves inferring a person's own mental state (Baron-Cohen 2000), while second-order theory of mind involves mental states about other peoples' mental states (Baron-Cohen 2000). While most children with HFA or Asperger syndrome are impaired in first and/or second order theory of mind functioning, most adults with HFA or Asperger syndrome show no impairment (Baron-Cohen 2000; Bowler 1992; Happé 1994; Ozonoff

et al. 1991a). This does not imply, however, that they are able to function adequately in social situations, since in daily life social information is more subtle and difficult to interpret (Ozonoff et al. 1991b). Therefore, 'advanced theory of mind' has been proposed as a more difficult level in theory of mind functioning compared to the first and the second level. Advanced theory of mind involves interpreting complex social situations, based on subtle information.

The most commonly used instruments to assess advanced theory of mind in high functioning adults with ASD are the 'Reading the mind in the eyes' test (further denoted as the 'Eyes test', Baron-Cohen et al. 1997b), the Strange stories test (Happé 1994) and the Faux-pas recognition test (Stone et al. 1998). Yet, only few studies exist that formally investigated these instruments in adults with HFA or Asperger syndrome and the results of these studies are mixed. While most studies reported impairments in advanced theory of mind in adults with HFA or Asperger syndrome (Baron-Cohen et al. 1997a, b, 2001; Happé 1994; Kaland et al. 2002; Stone et al. 1998; Zalla et al. 2008), two studies could not replicate these findings (Ponnet et al. 2004; Roeyers et al. 2001). These contradictory results may be attributed to the small research groups, which comprised at most 17 individuals with ASD, and to variations in the tests that were used. In the present study, we aim to investigate theory of mind in adults with ASD more thoroughly, by examining all three above-mentioned theory of mind tests in 61 participants with ASD, 32 of whom were diagnosed with HFA and 29 with Asperger syndrome. The results of the three tests will be compared with the performance of a matched neurotypical group.

A recent development in autism research is the use of self-reports to examine theory of mind functioning. Results showed that adults with ASD report impairment in their theory of mind abilities (Baron-Cohen and Wheelwright 2004). The use of self-reports in individuals with autistic impairment is controversial because of their hypothesized impaired introspective abilities (Frith and Happé 1999; Hobson 2005). However, as Frith and Happé (1999) describe, the 'explicit' theory of mind that characterizes adults with HFA or Asperger syndrome may enable them to recognize and describe their strengths and weaknesses adequately. Nevertheless, the results of self-reports to measure theory of mind have never been correlated with the results of neuropsychological tests to determine whether both point towards a similar phenomenon. The present study aims to fill this gap by using self-report questionnaires alongside neuropsychological tests to examine theory of mind and to investigate the relationships between these instruments.

In theory of mind research, it may be relevant to differentiate between HFA and Asperger syndrome. Although it is questionable whether HFA and Asperger syndrome can be differentiated, many researchers argue that these two disorders at least differ in degree of impairment (Klin et al. 2005; Ozonoff et al. 2000). In the present study, we differentiate between the two groups and match for verbal ability, since Frith (2004) suggested that differences in theory of mind performance between HFA and Asperger syndrome may reflect differences in verbal ability.

Hypotheses of the Present Study

We expect adults with HFA and the Asperger syndrome to be impaired, compared to neurotypical adults, in their performance of the three neuropsychological tests that measure theory of mind. We also expect them to report more theory of mind problems, resulting in lower scores on the EQ. Furthermore, as the neuropsychological tests and the self-report questionnaires measure comparable phenomena, medium to high correlations are expected between the tests and the self-report questionnaires.

## Methods

### Procedure

The participants of the HFA and the Asperger groups were recruited from GGZ (mental health institution) Eindhoven and GGZ Oost-Brabant. The participants visited one of these mental health institutions for various reasons, for example problems at work and/or marital problems. The recruitment took place from July 2005 to June 2008.

Participants with genetic conditions or relevant neurological, psychiatric or medical conditions (e.g. ADHD, Tourette syndrome) were excluded. Institutionalized patients were not included in order to ensure a relatively homogenous disorder group with relatively high functioning individuals.

Furthermore, the participants were selected for having at least a normal intelligence and verbal ability (scoring 85 or more in full scale intelligence and in the verbal comprehension index) as measured by the WAIS-III (Wechsler 1997). The neurotypical control participants were recruited from the general population by adds in local newspapers and by word of mouth. Typical controls were not included in the present study if they had a history of psychiatric illness or if autism ran in the family. In total, 93 of the 95 possible participants agreed to take part and signed informed consent forms prior to their inclusion in the present study. The individuals ranged in age from 18 to 60 years. The group comprised 32 individuals with HFA, 29 individuals with Asperger syndrome and 32 neurotypical adult controls. The mean Full Scale IQ of the

**Table 1** Matching variables

|  | HFA | Asperger | Neurotypical | Statistic | $p$-Value |
|---|---|---|---|---|---|
| Gender (male:female) | 32 (27:5) | 29 (25:4) | 32 (24:8) | $\chi^2 = 1.509$ | .47 |
| Mean age | 42.1 (10.8) | 43.67 (10.5) | 38.68 (9.3) | $F(2,90) = 1.92$ | .15 |
| WAIS scores |  |  |  |  |  |
| Full scale intelligence | 110.2 (13.8) | 114.5 (16.7) | 115.9 (10.0) | $F(2,90) = 1.47$ | .24 |
| Verbal comprehension | 110.8 (10.4) | 109.3 (12.6) | 113.9 (11.7) | $F(2,90) = 1.29$ | .28 |
| Perceptual organization | 105.6 (15.2) | 115.6 (15.3) | 114.0 (9.5) | $F(2,90) = 4.84$ | .01 |
| Freedom from distr. | 109.2 (16.0) | 108.8 (15.4) | 112.3 (11.6) | $F(2,90) = .56$ | .57 |
| Processing speed | 103.7 (19.4) | 110.8 (17.4) | 111.9 (14.6) | $F(2,90) = 2.10$ | .13 |

participants with HFA and Asperger syndrome and the neurotypical group was 110.2, 114.5 and 115.9, respectively (see Table 1). The present study was approved by the Ethics Committees of the two participating centers.

### Assessment of Disorder

The diagnosis of either HFA or Asperger syndrome was established through evaluation of historic and current symptomatology. To gather developmental information, parents were interviewed using the Dutch version of the Autism Diagnostic Interview, Revised version (ADI-R, Lord et al. 1994). When parental information was not available, an older brother or sister was interviewed. In these instances, further information about early childhood was gathered, for example from baby books and early clinical reports, until sufficient information was collected to fill in the diagnostic algorithm. The ADI-R was administered by psychologists who were officially trained in the administration and scoring of this instrument. Although the ADI-R has been validated only for children and adolescents, it is considered as the 'gold standard' for diagnosis, not only of children but also of adults (Lord and Corsello 2005).

In the process of diagnosing ASD, the ADI-R is often used in combination with the Autism Diagnostic Observation Schedule (ADOS, Lord et al. 1999). Research shows, however, that the ADOS is under-inclusive in diagnosing mild, verbal adolescents and adults with autistic spectrum disorders (Lord et al. 2000). Therefore, in the present study, observations of the participant were systematically gathered during the diagnostic process and during the assessment of the neuropsychological tasks. These observations were subsequently arranged according to the DSM-IV-TR criteria of ASD (American Psychiatric Association (APA) 2000). Furthermore, a semi-structured interview was administered to all participants, whereby all ASD criteria of the DSM-IV-TR were examined by asking the participant standard questions.

After the above diagnostic process, the DSM-IV-TR items of ASD were scored, based on the semi-structured interview and the observations of the participant. Only those participants who met the DSM-IV-TR criteria of the autistic disorder or Asperger syndrome were included in the present study. Because of the controversial nature of the DSM-IV criteria in differentiating between the two disorders (Ghaziuddin et al. 1992; Mayes et al. 2001), additional questions, based on the diagnostic criteria of Gillberg and Gillberg were asked (1989) and ICD-10 (World Health Organization 1993). When a significant delay in spoken or receptive language or development was present, a diagnosis of Asperger syndrome was excluded, following ICD-10 criteria. When there was no delay in development or language, the criteria of Gillberg and Gillberg (1989) were used to diagnose the participants with Asperger syndrome, since these criteria more closely resemble Asperger's own descriptions than the criteria of ICD-10 (Leekam et al. 2000).

### Assessment of Theory of Mind

To assess theory of mind, three neuropsychological tasks and one questionnaire were used. The participants were tested alone in a room that was free from distractions. The four theory mind tests were presented on paper and, in case of the Strange stories and the Faux pas test, they were read out by the experimenter. The tests were translated using a backward-forward procedure, in which the test was translated from English to Dutch and, subsequently, from Dutch to English by a second translator. Differences between translations were discussed, leading to a final translation.

More information about the tests used in the present study is described in the following paragraph.

### 'Reading the Mind in the Eyes' Test

The Eyes test was developed (Baron-Cohen et al. 1997a, b) and revised (Baron-Cohen et al. 2001) to measure subtle individual differences in social sensitivity of adults. We used the revised version of this Eyes test, which consists of 36 photographs of the region around the eyes of males and

females (Baron-Cohen et al. 2001). Participants have to decide which of four words best describes what the person in each photograph is thinking or feeling. In the present study, a Dutch translation of the test was administered. Translations were made according to a forward–backward procedure. The number of errors made by the participants was used as a measure of theory of mind functioning in the present study.

## Strange Stories Test

The Strange stories test was developed by Happé (1994) to measure advanced theory of mind. In the test, twenty-four vignettes present everyday situations in which people say things they do not mean literally. The stories were read aloud to the participants and the text of each vignette was placed in front of the participants, so they could also read the story themselves. Hereby the demands on working memory were reduced. The participants were asked questions about the intentions of the people in the vignettes. The eight stories that we chose were the most difficult for adults, as shown by studies of Happé (1994) and Jolliffe and Baron-Cohen (1999). These stories included misunderstanding, double bluff, irony, persuasion and white lies. Two scores were derived from the answers: First, the 'correct answer score', which is the sum of the scores for the answers about the intentions of the people in the stories (2 points for a fully correct answer, 1 point for a partially correct answer and 0 points for an incorrect justification; Happé et al. 1998). The second score was the number of stories for which the participant used a mental justification in their answers (in stead of a physical justification). All stories were scored by a second rater who was not involved in the testing process and who was unaware of the diagnostic status of the participants. The degree of concordance was 97% for the 'correct answer score' and 95% for the 'mental explanation score'. The test was officially translated into Dutch using a forward–backward procedure.

## Faux-Pas Test

In the Faux-pas test, participants were asked whether anyone in the story said something awkward and questioned the underlying motive. The experimenter read out each story, while the stories were placed in front of the participants so they could read the stories as well. After each story, questions were asked about the detection of the faux-pas (did anyone say something awkward?), about the person identification (who said something awkward?), about the content (what was awkward?), about the explanation (why was it awkward?), about the false belief (Did they know/remember that.) and an empathy question was asked (How did … feel?). The adult version of the Faux-

pas test was developed by Stone et al. (1998) and is based roughly on the children's version of the Faux-pas test (Baron-Cohen et al. 1999). A forward–backward procedure was followed for translation of the stories. The test comprises twenty stories, ten with and ten without a faux-pas. In the present study, four stories of both categories were randomly selected. The correct answer score was used as a variable in the present study. To score and interpret the answers, the instructions of Stone et al. (1998) were used. To validate the scoring procedure, the answers were also scored by a second rater. The degree of concordance of the total score was 95%.

## Empathy Quotient

The EQ is a self-report questionnaire, developed to examine empathizing tendencies in adults with normal intelligence (Baron-Cohen and Wheelwright 2004). Empathizing involves two elements: The ability to attribute mental states to oneself and others and to show an emotional reaction that is appropriate to the other person's mental state. In this definition, empathizing corresponds to what is meant by the term theory of mind (Baron-Cohen and Wheelwright 2004). The instrument comprises 60 questions, 20 of which are filler items and 40 items examine empathizing. The EQ proved to be a valid and reliable instrument (Lawrence et al. 2004). In the present study, a Dutch translation of the questionnaire was used. All participants filled in the EQ prior to receiving the results of their diagnostic process.

## Assessment of Intelligence

As part of the present study, the intelligence profile was assessed, using the Dutch version of the WAIS-III (Wechsler 1997). Four factors can be derived from WAIS-III data: Verbal Comprehension, Perceptual Organization, Freedom from Distractibility and Processing Speed. Factor analytic studies indicate that the four factor scales give the best estimates of the factors underlying intelligence (Arnou and Thompson 2000; Ryan and Paolo 2001). The norms have been improved to correct for the Flynn-effect that appeared to be present in the Dutch translation of WAIS-R. WAIS-III has excellent psychometric properties (Sattler and Ryan 1999) and has been validated for the Dutch population (Wechsler 1997).

## Matching Procedure

The three groups were matched according to age, gender, verbal abilities and Full Scale Intelligence Quotient (FSIQ). To match for verbal abilities, the WAIS-III factor scale 'Verbal Comprehension Index' (VCI) was used. The

subject characteristics of the three groups are presented in Table 1. The table shows that the three groups are well matched on nearly all characteristics. However, a significant difference was found with regard to the factor scale 'Perceptual organization' of the WAIS-III. The possible influence of this factor scale on theory of mind performance will be corrected for by using this subtest as a covariate (Field 2005).

## Results

### Differences in the Neuropsychological Tasks Measuring Theory of Mind

The mean scores and standard deviations of the theory of mind tests used in the present study are presented in Table 2.

To test the hypotheses of differences in the Eyes test, the Strange stories test and the Faux-pas test, three-one-way between-group multivariate analyses of covariance (MANCOVA) were performed. In each analysis, the diagnosis was used as the independent variable and the three neuropsychological tests as the dependent variables, respectively. The factor scale 'Perceptual organization' was used as a covariate to rule out that possible differences can be attributed to the differences in Perceptual organization between the three groups. The assumptions of homogeneity were met, however, Levene's test indicated that the assumption of equality of variance was not met in the analysis of the correct answer score and the mental explanation score of the Strange stories test. Therefore a more conservative alpha of .025 was set for these two variables (Tabachnick and Fidell 2001). After adjustment for Perceptual organization, a significant interaction effect was found with the correct answer score of the Strange stories test ($F$ (90) = 8.962, $p < .01$). With regard to the Eyes test and the Faux-pas test, no significant interaction effects of Perceptual organization were found.

For the Eyes test, no statistically significant main effect of diagnosis was found ($F$ (2,90) = .959, $p = 0.39$). For the correct answer score of the Strange stories test, the results did reveal a statistically significant main effect of diagnosis ($F$ (2,90) = 7.570, $p < .01$ partial eta squared = .09), which can be described as moderate (Cohen 1988). Post-hoc Tukey comparisons showed that the HFA group ($p = .03$) and the Asperger syndrome group ($p < .01$) made significantly more errors compared to the neurotypical control group. No significant main effect was found for the mental explanation score in the Strange stories test ($F$ (2,90) = 1.019, $p = .82$).

In performance of the Faux-pas test, a main effect of diagnosis appeared ($F$ (2,90) = 6.984, $p < .01$ partial eta squared = .14), with a large effect size (Cohen 1988). Post-hoc Tukey comparisons indicated that the adults with Asperger syndrome had a significantly ($p < .01$) lower total score compared to the neurotypical adults. A trend toward an effect was visible for the differences between the HFA group and the neurotypical group ($p = .06$).

No differences were found between the HFA and the Asperger syndrome group in the Eyes test, the Strange stories test or the Faux-pas test.

### Differences in Self-reported Theory of Mind

The mean scores and standard deviations of the EQ for the HFA group, the Asperger syndrome group and the neurotypical group are presented in Table 2. To examine the hypothesis of differences in self-reported theory of mind, a one-way between-group multivariate analysis of variance was performed with the diagnosis as the independent variable or factor and the EQ as the dependent variable. Again,

**Table 2** Means and SD for the tests

| | $M$ | SD | $M$ | SD | $n$ |
|---|---|---|---|---|---|
| | *Eyes test: errors made* | | *Faux-pas test: total score* | | |
| HFA | 12.78 | 4.81 | 20.28 | 3.40 | 32 |
| Asperger | 11.86 | 3.88 | 18.97 | 3.95 | 29 |
| Neurotypical group | 11.00 | 3.59 | 22.22 | 2.70 | 32 |
| | *Strange stories: correct answers* | | *Strange stories: mental answers* | | |
| HFA | 14.13 | 1.96 | 7.88 | .34 | 32 |
| Asperger | 13.62 | 2.29 | 7.79 | .56 | 29 |
| Neurotypical group | 15.31 | 1.09 | 7.94 | .25 | 32 |
| | *Total score empathy quotient* | | | | |
| HFA | 25.41 | 11.38 | | | 32 |
| Asperger | 24.66 | 9.94 | | | 29 |
| Neurotypical group | 51.56 | 11.53 | | | 32 |

the factor scale 'Perceptual Organization' was used as covariate. The analyses showed that the assumptions of homogeneity and equality of variance were met. Wilks' Lambda was used to measure group differences. No interaction effect was found between Perceptual Organization and the EQ score ($F$ (2,90) = .662, $p$ = .42). For the EQ score, a statistically significant main effect of diagnosis was found ($F$ (2,90) = 58.938, $p$ < .01, partial eta squared = .57), with an effect size that can be interpreted as very large (Cohen 1988). To investigate which differences between the three diagnostic groups added to the main effects, post-hoc Tukey comparisons were performed. Analyses showed that the neurotypical group yielded significantly higher EQ-scores in comparison to the HFA ($p$ < .01) and the Asperger syndrome group ($p$ < .01). The scores of the two disorder groups did not differ significantly. The findings support the hypothesis postulated in the present study of impaired theory of mind in the HFA and Asperger syndrome groups.

## The Association between the Neuropsychological Tasks and the Self-report Questionnaire

To test the hypothesis that the total score for the self-report questionnaire is closely related to performance of the neuropsychological tasks measuring theory of mind, Pearson product-moment correlation coefficients were calculated for the total group. Preliminary analyses were performed to ensure that assumptions of normality, linearity and homoscedasticity were not violated. Only medium or high correlations that reached significance will be described. Table 3 shows the correlation matrix for these results.

Medium-sized significant correlations were found between the correct answer score of the Strange stories test and the Faux-pas test ($r$ = .36, $p$ < .001), between the EQ and the Faux-pas test ($r$ = .31, $p$ < .005) and between the correct answer score of the Strange stories test and the EQ ($r$ = .29, $p$ < .005).

## The Ability of the Tests to Predict Whether a Diagnosis is Present

Because we did not find any differences between the two disorder groups in the previous analyses, we decided to merge these two groups into one diagnostic group for further analysis.

In order to examine the ability of the tests to predict whether a certain person belonged to the diagnostic group or the neurotypical group, we performed a logistic regression analysis. The presence of a diagnosis was entered as the dependent variable and the three tasks and the self-report questionnaire were entered as the independent variables. The Maximum Likelihood model was used to estimate the parameters. The Goodness of Fit of the analysis as measured by the Omnibus Tests of Model Coefficients and the Hosmer and Lemeshow test was sufficient. The results of the logistic regression analysis are presented in Table 4. To determine the predictive power of the model the Classification results were calculated. Table 5 shows the results.

Table 5 shows that the percentages of correct classifications are high: 95% are correctly placed in the disorder group and 88% are correctly placed in the neurotypical group, correspondingly implying low numbers of false positives (12%) and false negatives (5%).

Table 4 illustrates that only the EQ measure is predictive for whether a participant belonged to the neurotypical group or the diagnostic group when the significance level is set at alpha = 0.01. This table further shows that the correct answer score of the Strange stories test ($p$ = .02) and the Faux-pas test ($p$ = .05) are also predictive when the significance level is set at alpha = 0.05. Results of the Eyes test and the mental answer score of the Strange stories test were not significant. These findings suggest that the Strange stories test and the Faux-pas test are valuable instruments for examining theory of mind in adults when self-reports cannot be used.

Since the EQ was the most predictive for the presence of a diagnosis, we further examined the sensitivity of the EQ

**Table 3** Correlation coefficients

|  | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Total group |  |  |  |  |  |
| Reading the eyes test | – |  |  |  |  |
| Faux-pas test | −.181 | – |  |  |  |
| Strange stories: correct score | −.226* | .359** | – |  |  |
| Strange stories: mental answers | .024 | .164 | .387** | – |  |
| Empathy quotient | −.213* | .305** | .294** | .094 | – |

\* $p$ < .05; \*\* $p$ < .01

**Table 4** Logistic regression analysis

|  | $B$ | SE | $p$-Value | Wald | OR |
|---|---|---|---|---|---|
| Logistic regression analysis neuropsychological tests and self-report |  |  |  |  |  |
| Faux-pas test: total score | .201 | .117 | .09 | 2.926 | 1.222 |
| Strange stories test: correct | .332 | .279 | .23 | 5.133 | 1.394 |
| Strange stories test: mental | .598 | 1.411 | .67 | .179 | 1.818 |
| Reading the mind in the eyes | .024 | .111 | .83 | .046 | 1.024 |
| EQ | .169 | .036 | .00 | 22.160 | 1.184 |

**Table 5** Classification table

| Observed | Neurotypical group versus disorder group | |
|---|---|---|
| | Disorder group | Neurotypical group |
| Disorder group | 58 (95%) | 3 (5%)[a] |
| Neurotypical group | 4 (12%)[b] | 28 (88%) |

[a] False positives

[b] False negatives
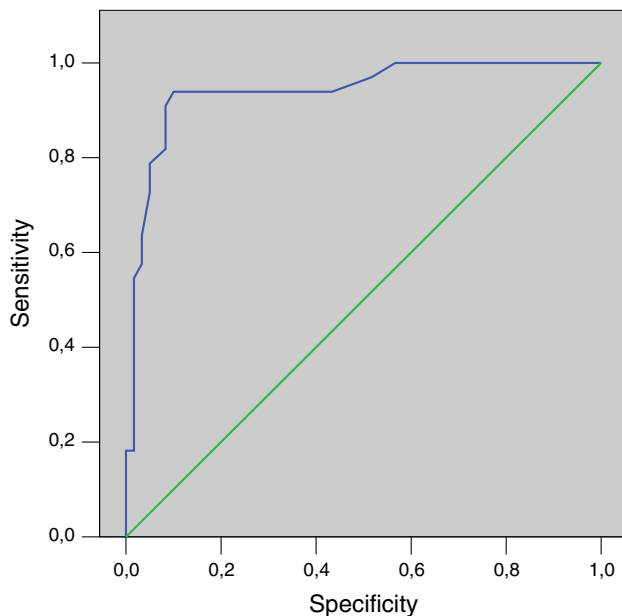


**Fig. 1** Roc curve EQ

in making a correct group classification by calculating a Receiver Operating Characteristic analysis (ROC-analysis; Fawcett 2006). The ROC curve is illustrated in Fig. 1.

The analysis showed that the participants could be assigned correctly to either the neurotypical or the HFA/Asperger syndrome group in 94% of the cases, based on their EQ scores. This suggests that the EQ is a highly sensitive marker of the presence of a diagnosis (HFA or Asperger syndrome).

**Discussion**

The present study aimed to assess theory of mind functioning in adults with HFA or Asperger syndrome and a neurotypical group. Differences were found for the Faux-pas test and the Strange stories test in the expected direction, indicating theory of mind impairment in the two disorder groups. These findings are in line with the results of the EQ, which showed more self-reported theory of mind problems in the HFA and Asperger syndrome groups compared to the neurotypical individuals. In contrast to our

expectations, performance of the Eyes test did not differ between the three groups. Despite, since impairment was found for the disorder groups by three of the four theory of mind tests, the hypothesis of theory of mind impairment in adults with HFA or Asperger syndrome is confirmed.

As far as the relationships between the neuropsychological tests and self-reports are concerned, the analyses showed medium and significant correlations between the EQ, the Strange stories test and the Faux-pas test. This suggests that these three tests measure a similar underlying cognitive phenomenon, which is also in line with our expectations.

No differences were found between the HFA and the Asperger syndrome group for any of the tests that were used. These results replicate previous theory of mind research in children (Dahlgren and Trillingsgaard 1996). Apparently, when corrected for verbal abilities, theory of mind ability is similar in adults with HFA and Asperger syndrome. This is at odds with the hypothesis that these two groups differ in degree of impairment (Klin et al. 2005; Ozonoff et al. 2000). Our findings do confirm other studies that question the validity of distinguishing Asperger syndrome and HFA as separate entities (Volkmar and Klin 2005).

The results that were found for each of the four theory of mind tests will be specifically discussed in the following paragraphs.

The Eyes test is the only neuropsychological test in the present study which revealed no impairment for the HFA or the Asperger syndrome group. Whilst our results replicate the results of Roeyers et al. (2001), they are at odds with findings of the original Eyes test (Baron-Cohen et al. 1997a, b) and of the Revised Eyes test (Baron-Cohen et al. 2001). These contradictory results cannot be attributed to inaccuracy of the translation of the test, since was done carefully by following a forward–backward procedure.

When taking a closer look at our results, it is remarkable that the relationships between the Eyes test and the two other neuropsychological tests that measure theory of mind are weak or even absent. Furthermore, in contrast to the Strange stories test and the Faux-pas test, the relationship with the self-reported theory of mind is weak. This underlines the hypothesis, put forward by Roeyers et al. (2001), that the Eyes test may not be a valid measure of advanced theory of mind. We need to stress here, however, that although the translation of the test was done carefully, cultural aspects may have played an important role. In each culture, there are implicit guidelines regarding the appropriateness of showing certain facial expressions in situations, which are called 'display rules' (Ekman and Friesen 1969). Those implicit rules may be different for Dutch people compared to the English population that was investigated in the study of Baron-Cohen et al. (2001), leading to different scores on the Eyes test in a Dutch

population. Interestingly, Roeyers et al. (2001) did not use the original version of the original Eyes test. Instead, they constructed a new Dutch version of the test, which makes the influence of display rules on performance unlikely. However, their results were similar to our results.

Another relevant factor in the performance on the Eyes test, may be the difficulty that was experienced by the neurotypical group. They frequently expressed the need for more information than only the eyes in order to correctly identify the emotions. This may be attributed to the drive for 'central coherence' that characterizes most neurotypical individuals (Frith 1989). The tendency to integrate information in the context may be responsible for the relatively high mean error rate for the neurotypical group and the lack of differences between the three groups. Summarizing, it is questionable whether the Eyes test is a valid indicator of theory of mind in high-functioning adults with ASD, at least in a Dutch population. Performance of this task probably reflects other cognitive domains.

The strength of the Strange stories test as opposed to the Eyes test is that it closely resembles social situations as they occur in daily life. Our data showed that nonverbal reasoning skills influenced performance on this task. Apparently, the ability to analyze the (relevant variables in the) stories influenced the results. However, after correcting for nonverbal reasoning skills, the difference between the disorder groups and the neurotypical group still remained significant. Our results replicate the findings of previous studies of adolescents and adults with Asperger syndrome (Jolliffe and Baron-Cohen 1999; Kaland et al. 2002) or HFA (Happé 1994; Jolliffe and Baron-Cohen 1999). Contradictory results were reported by Roeyers et al. (2001) who found no impairments in performance of the Strange stories test for a group of adolescents and adults with ASD. However, Roeyers et al. (2001) did not match their research groups with regard to verbal ability and age, while verbal abilities especially have been found of distinct influence on this verbal theory of mind test (Happé 1994). The present study did show that a considerable proportion of the participants with HFA and Asperger syndrome performed faultlessly. For at least some of the individuals in the HFA and the Asperger syndrome groups, the Strange stories test may have been too easy. This is not surprising since the test was originally developed for children, while our group consisted of relatively able adults. Although the most complicated stories were chosen for the present study, the level of difficulty is considerably lower than in real life social situations. This probably also explains why no differences were found between HFA and Asperger syndrome in the use of mental explanations in the Strange stories test, confirming previous findings of Happé (1994) and Roeyers et al. (2001) in adults with ASD.

Contrary to the Strange stories test, the Faux-pas test has been specifically developed for adults. The social scenarios in the vignettes closely resemble situations that occur on a regular basis in daily life (Baron-Cohen et al. 1999; Stone et al. 1998). Therefore, the ecological validity of this test is expected to be high, which means that performance on the Faux-pas test gives a valid indication of everyday cognitive ability (Chaytor et al. 2006). In the present study, the Asperger syndrome group was significantly impaired in performance of this test and impairment in the HFA group was near significant. These results confirm recent findings of Zalla et al. (2008) for adults with Asperger syndrome. Given the theory of mind impairment in individuals with ASD and considering the impairments found with the Strange stories test, we expected larger differences between the two disorder groups and the neurotypical group for this test. After taking a closer look at the results of the Faux-pas test, it became apparent that, especially compared to the Strange stories test, the neurotypical individuals made relatively many errors. Apparently, this specific test is also difficult for neurotypical adults.

In summary, we propose that the Faux-pas test and the Strange stories test are valuable instruments for clinical use because they closely resemble everyday social situations. In case of the Faux-pas task it is important to take into account the fact that neurotypical individuals usually do not perform faultlessly and that error rates need to be relatively high in order to provide a valid indication of theory of mind impairment. As for the Strange stories test, low error rates may be indicative for theory of mind impairment, whilst a faultless performance does not rule out subtle impairment in theory of mind.

The present study also examined self-reported theory of mind using self reported information by examining the EQ. Large differences were found between the neurotypical group and the two disorder groups, which agrees with previous results of Baron-Cohen and Wheelwright (2004). The correlations between the EQ, the Strange stories and the Faux-pas test confirm the hypothesis that the two disorder groups are able to recognize their theory of mind abilities adequately. This is in line with previous research, which demonstrated that high functioning adults with ASD have more self-knowledge and introspective abilities than was previously assumed (Blackshaw et al. 2001; Frith and Happé 1999; Happé 1991; Spek et al., in preparation). Somewhat similar results have also been found for adults with schizophrenia, were EQ performance appeared modestly associated with social cognitive tasks (Bora et al. 2008).

The present study thus lends support to the validity of self-reports for examining theory of mind in adults with HFA or Asperger syndrome, not only in research but also in clinical practice. The large discriminating power of the

EQ in differentiating between the two disorder groups and the neurotypical group indicates that the total score of the EQ may be considered an important marker for the presence of HFA or Asperger syndrome in adults.

Limitations

The present study was undertaken in adults with average to high verbal abilities. Therefore, the results deriving from this study cannot be generalized to ASD populations with below average verbal abilities.

Adequate understanding and interpretation of the questions used in the EQ relies on semantic capacities. Although the two disorder groups were carefully selected and all participants had at least average verbal abilities, deficiencies in semantic processing which characterize individuals with ASD may have influenced the answers to the questions.

The present study indicates that performance of the EQ may function as a marker for the presence of HFA or Asperger syndrome, when compared to a neurotypical group. However, to be a clinical marker it is of great importance that the EQ can also differentiate between ASD and other psychiatric diagnoses. Further research should shed more light on this.

## References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, text revision* (4th ed.). Washington, DC: American Psychiatric Association.

Arnou, R. C., & Thompson, B. (2000). Second order confirmatory factor analysis of the WAIS-III. *Assessment, 7*(3), 237–246.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds. Perspectives from developmental cognitive neuroscience* (pp. 3–21). Oxford: University Press.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163–175.

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., Jolliffe, T., Mortimore, C., & Robertson, M. (1997a). 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., Wheelwright, S., & Jolliffe, T. (1997b). 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., O'Riordan, M., Jones, R., Stone, V. E., & Plaisted, K. (1999). A new test of social sensitivity. Detection of faux-pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders, 29*, 407–418.

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*, 241–252.

Blackshaw, A. J., Kinderman, P., Hare, D. J., & Hatton, C. (2001). Theory of mind, causal attribution and paranoia in Asperger syndrome. *Autism, 5*(2), 147–163.

Bora, E., Gökçen, S., & Veznedaroglu, B. (2008). Empathic abilities in people with schizophrenia. *Psychiatry Research, 160*, 23–29.

Bowler, D. M. (1992). Theory of mind in Asperger syndrome. *Journal of Child Psychology and Psychiatry, 33*, 877–895.

Chaytor, N., Schimitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*, 217–227.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.

Dahlgren, O., & Trillingsgaard, A. (1996). Theory of mind in non-retarded children with autism and Asperger's syndrome. A research note. *Journal of Child Psychology and Psychiatry, 37*(6), 759–763.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica, 1*, 49–98.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Field, A. (2005). *Discovering statistics using SPSS*. California: Sage.

Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford: Blackwell.

Frith, U. (2003). *Autism: Explaining the Enigma*. Oxford: Blackwell.

Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry, 45*(4), 672.

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

Ghaziuddin, M., Tsai, L. Y., & Ghaziuddin, N. (1992). A comparison of the diagnostic criteria for Asperger syndrome. *Journal of Autism and Developmental Disorders, 22*, 634–639.

Gillberg, I. C., & Gillberg, C. (1989). Asperger syndrome—some epidemiological considerations: a research note. *Journal of Child Psychology and Psychiatry, 30*, 631–638.

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

Happé, F. (1994). An advanced test of theory of mind: understanding of story character 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.

Hobson, P. (2005). Autism and emotion. In F. R. Volkmar, P. Rhea, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 406–422). New Jersey: Wiley.

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, 43*(4), 517–528.

Klin, A., McPartland, J., & Volkmar, F. R. (2005). Asperger syndrome. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 88–125). New Jersey: Wiley.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine, 34*, 911–924.

Leekam, S., Libby, S., Wing, L., Gould, J., & Gillberg, C. (2000). Comparison of ICD-10 and Gillberg's criteria for Asperger syndrome. *Autism, 4*(1), 11–28.

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

Lord, C., & Corsello, C. (2005). Diagnostic instruments in autistic spectrum disorders. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 730–772). New Jersey: Wiley.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

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

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism diagnostic observation schedule—WPS (ADOS-WPS)*. Los Angeles, CA: Western Psychological Services.

Mayes, S. D., Calhoun, S. L., & Crites, D. L. (2001). Does DSM-IV Asperger's disorder exist? *Behavioral Science, 29*(3), 263–271.

Ozonoff, S., Pennington, B. F., & Rogers, S. (1991a). Executive function deficits in high- functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*(7), 1081–1105.

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

Ozonoff, S., South, M., & Miller, J. N. (2000). The external validity of Asperger disorder: Lack of evidence from the domain of neuropsychology. *Journal of Abnormal Psychology, 109*, 227–238.

Ozonoff, S., South, M., & Provencal, S. (2005). Executive Functions. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 606–627). New Jersey: Wiley.

Ponnet, K. S., Roeyers, H., Buysse, A., De Clercq, A., & Van der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism, 8*(3), 249–266.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

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.

Rumsey, J. (1985). Conceptual problem solving in highly-verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders, 15*, 23–36.

Ryan, J. J., & Paolo, A. M. (2001). Exploratory factor analysis of the WAIS-III in a mixed patient sample. *Archives of Clinical Neuropsychology, 16*(2), 151–156.

Sattler, J. M., & Ryan, J. J. (1999). *Assessment of children, revised: WAIS III supplement*. LeMesa, CA: Jerome Sattler Publishing.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*, 640–656.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Volkmar, F. R., & Klin, A. (2005). Issues in the classification of autism and related conditions. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 606–627). New Jersey: Wiley.

Wechsler, D. (1997). *Manual for the Wechsler adult intelligence scale—Third edition (WAISI-III)*. San Antonio, TX: Psychological Corporation.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioral disorders. Diagnostic criteria for research*. Geneva: WHO.

Zalla, T., Sav, A., Stopin, A., Ahade, S., & Leboyer, M. (2008). Faux-pas detection and intentional action in Asperger syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorders, 93*, 373–382.