

Relationship between Standard False Belief, Nonvocal False Belief and Guesser-Knower Tests
in Children with Autism Spectrum Disorders

Kerri Walters

University of Manitoba

Masters Thesis

Submitted to the Department of Psychology

in partial fulfillment of the Master of Arts Degree at the University of Manitoba

2006

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

**Relationship between Standard False Belief, Nonvocal False Belief and Guesser-Knower Tests in
Children with Autism Spectrum Disorders**

by

Kerri Walters

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of

Manitoba in partial fulfillment of the requirement of the degree

of

Master of Arts

Kerri Walters © 2006

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Acknowledgments

Appreciation extends to families and children for their participation throughout this study. Further acknowledgements are extended to Dr. C.T. Yu for his guidance and mentorship throughout this research; to Drs. Garry Martin, Angela Cornick, and Dennis Hrycaiko, my committee members, for their helpful comments and support; to May Lee for her assistance in statistical analyses; and to Rossana Astacio, Melissa Lam, Carole Marion, Jennifer Thorsteinsson, and Duong Nguyen for their assistance with the research.

This research was supported by the University of Manitoba Graduate Fellowship, the Manitoba Graduate Scholarship, the Social Sciences and Humanities Research Council Masters Fellowship, the Shannon Hamm Award, and the Autism Spectrum Disorders Canadian-American Research Consortium Strategic Training Initiatives in Health Research grant which is jointly funded by the Canadian Institute of Health Research and the National Alliance for Autism Research.

List of Tables

<i>Table 1.</i>	Children's characteristics and test results.....	17
<i>Table 2.</i>	Percent correct responses to questions on false belief tests and percent of children who passed the test.	27

Table of Contents

Acknowledgments.....	2
List of Tables.....	3
1. Abstract.....	6
2. Introduction	
Autism Spectrum Disorders.....	7
Research on Perspective Taking.....	8
FB test involving false location	10
FB test involving false content.....	10
Research on Language and Perspective Taking.....	12
Research on Nonvocal Component Skills of Perspective Taking with Non-human Primates.....	13
Statement of the Problem.....	15
3. Method	
Participants and Setting.....	16
Apparatus.....	18
False location and false content standard FB tests.....	18
False location and false content nonvocal FB tests.....	19
Guesser-Knower test.....	19
Experimental Design.....	19
Standard False Belief Assessment Procedures.....	20
Nonvocal False Belief Assessment Procedures.....	21

Guesser-Knower Assessment Procedures.....	22
Goggles preexposure.....	22
Guesser-Knower assessment.....	23
Interobserver Reliability and Procedural Integrity Assessments.....	24
Interobserver reliability.....	24
Procedural integrity.....	25
4. Results.....	26
5. Discussion.....	29
6. References.....	33
7. Appendices	
Appendix A. Scripts for Standard False Belief Tests	38
Appendix B. Scripts for Nonvocal False Belief Tests.....	40
Appendix C. Procedural Checklist for Guesser-Knower Test.....	42

Abstract

Approximately 80% of children with autism spectrum disorders (ASDs) fail to perform perspective taking as measured by false belief (FB) tests that require speech. Since children with ASDs exhibit a range of verbal abilities, it is possible that children with limited speech may be successful on false belief tests that do not require vocal responses. The present study examined the performance of children on five false belief tests that require different verbal skills. Eighteen children with ASDs (ages 5 years to 11 years 11 months) were tested on: (a) two standard FB tests that require children to answer questions vocally; (b) two nonvocal FB tests that allow children to point to visually displayed options when responding to questions; and (c) a nonvocal Guesser-Knower (GK) test designed to evaluate whether children are able to discriminate the relationship between seeing and knowing. Results indicate that: (a) performance did not differ significantly among the five different tests; (b) children with higher expressive language scores outperformed those with lower expressive language scores on all tests; (c) children with Asperger syndrome outperformed children with autistic disorder on all five tests; and (d) children with higher working memory scores outperformed children with lower working memory scores on all five tests.

Relationship between Standard False Belief, Nonvocal False Belief and Guesser-Knower Tests in Children with Autism Spectrum Disorders

Children with autism spectrum disorders (ASDs) have been consistently shown to present difficulty in perspective taking. The inability to take another person's perspective has been considered a keystone deficit in reciprocal social interactions for this population. Perspective taking is measured by false belief (FB) tests that require a child to answer a series of questions about how another person understands a particular event based on that person's perspective. Given that 80% of children with ASDs fail FB tests, research on component skills of perspective taking is much needed. Considering that communication is one of the core deficits in the diagnosis of ASDs, this finding evokes the question: Can children with ASDs who have failed FB tests succeed on similar tests that do not require expressive language? Therefore, the present study examined the performance of children on standard FB tests and nonvocal FB tests. It has also been suggested that understanding the relationship between "seeing and knowing" is fundamental to perspective taking (Perner, Frith, Leslie, & Leekam, 1989). Therefore, children in this study were also assessed on a nonvocal test designed to evaluate this discrimination.

Autism Spectrum Disorders

Autism spectrum disorders or ASDs refer to a subgroup of pervasive developmental disorders (American Psychiatric Association, 2000) including autistic disorder, Asperger syndrome, and pervasive developmental disorder-not otherwise specified (PDD-NOS). They are characterized by severe impairment in social interactions and by restricted behaviour and interests (e.g., ritualistic and repetitive movements). Autistic disorder and PDD-NOS are further characterized by a deficit in communication skills including expressive and receptive language. Children who exhibit characteristics associated with autistic disorder, but who do not meet the

diagnostic criteria would be given a diagnosis of PDD-NOS.

Social skills impairment is a primary characteristic of ASDs and impairment in reciprocal social interaction is gross and sustained (Baron-Cohen, 1988; Fein, Pennington, Markowitz, Braverman, & Waterhouse, 1986; Rutter, 1983). Children with ASDs display a wide range of verbal skills from being mute to emitting speech that is comparable to their typical peers. The onset of ASDs occurs before 3 years of age and is diagnosed more frequently in boys (ratio of 4 boys to 1 girl; Pelios & Lund, 2001).

As one of the most common neurodevelopmental disorders, autistic disorder affects approximately 11 per 10,000 children and its prevalence has been increasing over the past 15 years. Based on 19 surveys conducted in 10 countries published between 1987 and 2001, Fombonne (2003) estimated that the prevalence of ASDs ranged from 2.5 to 30.8 per 10,000.

Research on Perspective Taking

The ability to take another person's perspective is impaired in individuals with ASDs (Baron-Cohen, Leslie, & Frith, 1985, 1986). Operationally, perspective taking is typically studied using structured tasks that require the child to: (a) distinguish his or her own perspective from that of others, and (b) predict the behaviour of the other person based on that person's perspective (e.g., an experimentally arranged false belief) as opposed to one's own.

Behaviourally, we might say that the child is required to: (a) discriminate between two sets of environmental stimuli (e.g., what the child sees and what another person can see, which may be different); and then (b) predict or describe how the other person would respond in a given situation.

Perspective taking can vary in complexity. Level 1 (or first order) perspective taking is defined as the ability to identify the perspective of another person about an objective event (e.g.,

what does Mary think?). Level 2 (or second order) perspective taking is the ability to identify what a second person would say about the perspective of a third person about an objective event (e.g., what does Mary think Jon thinks? Flavell, Botkin, Fry, Wright, & Jarvis, 1968). The order of complexity increases with each iteration. The focus of the present research was on Level 1 perspective taking.

A false belief is arranged by changing the circumstances while someone is unable to observe that change, thus resulting in that person's continued belief in the original circumstance. For example, suppose that Fred is on his way to work and suddenly realizes that he left his lunch bag in the refrigerator at home. He decides to turn around and head back home to get his lunch bag. In the meantime, Fred's wife found his lunch bag in the fridge and decided to bring it to his office. When Fred arrived at home he began searching for his lunch bag in the fridge. Although the circumstances (i.e., location of Fred's lunch bag) had changed he was not there to observe the change. Consequently, Fred maintained a "false belief" about the location of his lunch bag and behaved accordingly.

The two commonly used tests designed to evaluate Level 1 perspective taking require the child to predict the behaviour of another person given information about that person's false belief and the child's knowledge about the true state of affairs. To pass the FB tests the child is required to convey that a person's behaviour is determined by his/her beliefs about reality, rather than by reality itself, even when those beliefs are wrong (i.e., false). One type of assessment involves *false location* whereby the child must identify another person's false belief about the location of an item (Baron-Cohen et al., 1985). Another type of assessments involves *false content*, during which the child must identify the false belief another person has about the contents of a container (Perner et al., 1989).

FB test involving false location. In the "Sally Anne false location test" (Baron-Cohen et al., 1985), a child is in a room with two dolls named Sally and Anne (puppeteered by experimenters). Sally puts a marble in the basket and leaves the room. Anne then removes the marble from the basket and places it in a box located on the table in front of her. When Sally returns, the experimenter asks the child: (a) "where will Sally look for the marble?" the correct answer would be "in the basket"; (b) "where is the marble really?" the correct answer would be "in the box"; and (c) "where did Sally and Anne put the marble before Sally left the room?" the correct answer would be "in the basket". In order to pass this FB test, the child must correctly answer all three of these questions.

FB test involving false content. In the "M&M False Content test" (Perner et al., 1989), a child is shown a box of M&M's and is asked "what do you think is inside of this box?" the correct answer would be M&M's or candy. Following the child's response, the box is opened to reveal a pencil. The child is then asked what another child, who has not seen the contents, will think the box contains (i.e., "what will Jon say is inside of this box?"). Finally, the child is asked "do you remember when I first showed you this box. What did you say was inside of it?" If the child correctly responds to these questions, a pass is scored.

Both tests contain: (a) a prediction question requiring the child to indicate what another person would say about an event, (b) a reality question requiring the child to indicate the true state of events, and (c) a memory question requiring the child to recall the state of events before the false belief was introduced. Correct responding to these questions determines whether an individual is able to take the perspective of another person.

By the age of 2, typically developing children understand that other people have different desires from their own and that these desires are responsible for governing their behaviour

(Gopnick & Slaughter, 1991). By the age of 3, they understand that beliefs influence behaviours; after age 3, they understand that even when such beliefs are false, people will behave according to their beliefs (Siegler, 1986). Most typically developing children pass the above FB tests by age 4 (Baron-Cohen et al., 1985). On the other hand, approximately 80% of children with ASDs with a verbal age between 4 and 5 years could not pass the FB tests (Baron-Cohen et al., 1985, 1986; Perner et al., 1989; Swettenham, 1996). Children with ASDs either do not develop perspective taking or have a significant delay in acquiring this ability (Baron-Cohen et al., 1985). Non-intervention studies have demonstrated that it remains a persistent deficit that does not improve as mental age advances (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994). Baron-Cohen et al. (1985) examined the relationship between IQ and FB performance, in which 16 of 20 children diagnosed with autistic disorder between the ages of 6.1 and 16.6 years and whose IQ scores, based on non-verbal mental age, averaged 82 (range = 70 to 107) were unable to pass the Sally Anne FB test. In contrast, 85% of typically developing children between the ages of 3.5 and 5.9 years of age were able to pass the same assessment. Further, 86% of children diagnosed with Down syndrome between the ages of 6 years 3 months and 17 years of age whose IQ scores, based on verbal mental age, averaged 64 (range = 42 to 89) were able to pass the assessment. While support for a perspective taking deficit is clear in children with ASDs, few studies have examined the component skills that make up perspective taking.

To date, research tends to treat perspective taking as an all or none phenomenon, a fact which is evidenced by the frequent use of a dichotomous pass/fail scoring criterion exhibited in previous research based on correct responding to prediction, reality, and memory questions (e.g., Baron-Cohen et al. 1985; Charlop-Christy & Daneshvar, 2003). Other researchers have allotted partial points to the child for demonstrating just the prediction and reality components of the FB

test (e.g., Hogrefe, Wimmer, & Perner, 1986; Leslie & Frith, 1988; Muris et al., 1999; Perner et al., 1989). The inconsistency in pass/fail criteria among investigators suggests controversy and a lack of consensus on what prerequisite, partial, or component skills make up the ability to take someone's perspective.

Extending perspective taking assessments beyond the standard FB tests would have three important benefits. First, adherence to the standard FB tests could leave the impression that perspective taking might not develop beyond this level (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999). Baron-Cohen et al. (1999) proposed that success on these tests should be interpreted as relatively early points in the acquisition of perspective taking, rather than the end point. Second, for individuals who pass FB tests there is a ceiling effect, where no difference can be detected between high functioning children with ASDs and typically developing children. Evidence of a need to develop perspective taking assessments that extend beyond the standard FB tests lies in the subgroup of individuals with ASDs who pass the standard FB tests despite their self-reports of difficulty understanding other people's thoughts, intentions, and actions (Bowler, 1992; Dahlgren & Trillingsgaard, 1996; Happé, 1994). To address this, some researchers have developed more advanced tests of perspective taking in order to further examine individuals who pass standard FB assessments (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelright, & Jolliffe, 1997; Happé, 1994). Finally, standard FB tests are limited to children with speech. Examining the prerequisite, or partial, perspective taking skills for children without expressive language is equally important. Currently, the literature on prerequisite or composite skills and on perspective taking of nonvocal children with ASDs remains underdeveloped.

Research on Language and Perspective Taking

Research examining children with specific language impairments has focused primarily on children whose impairment involves language comprehension (e.g., Leslie & Frith, 1988; Perner, et al., 1989). The results showed that typically developing children with receptive language deficits consistently outperformed children with ASDs, suggesting that receptive language deficits are not responsible for failure to perform standard FB tests. However, research shows that children who lack expressive language are at a clear disadvantage when tested for perspective taking. When the children were taught language skills specific to the FB tests, their performance improved (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Olson, 1988). This is an important finding given that success on the standard FB tests is dependent on expressive language (e.g., when asked "What do you think Jon will say is inside the box?" the child must respond by saying "candy" or "M&Ms"). The results of previous research support the argument that language acquisition plays a role in standard FB test performance. Furthermore, research examining verbal mental age (VMA; which includes both expressive and receptive language) and performance on standard FB tests has demonstrated that VMA is highly correlated with performance on perspective taking tests among children with ASDs as well as children who were considered to be typically developing (Happé, 1995).

Research on Nonvocal Component Skills of Perspective Taking with Non-human Primates

Perspective taking has been examined in non-human primates through testing for an understanding of the relationship between seeing and knowing. Behaviourally, "seeing" is defined as the absence of a barrier between a person and a given object and "knowing" is defined as the stimulus control exerted by "seeing" on subsequent behaviours in relation to the object. Assuming no deception or errors, these "subsequent behaviours" should be consistent with what has or has not been seen. For example, a person who has *seen* an object being hidden should

correctly identify the location of the object 100% of the time when requested to do so, whereas a person who has *not seen* an object being hidden would not. The ability to discriminate between persons who have or have not seen a given event and to predict their subsequent behaviours appears to be a critical precursor or component skill for passing standard FB tests.

Povinelli, Nelson, and Boysen (1990) studied “seeing and knowing” in a two-stage procedure with chimpanzees. During the *first stage*, a chimpanzee was in a room with two trainers. One of the trainers, designated as the “Guesser”, left the room and the other, designated as the “Knower”, baited one of four containers. The containers were located behind a screen so that the chimpanzee could not see which container was baited, but could see who (the Knower) had done the baiting. Following baiting, the Guesser returned to the room, the screen was removed, and each trainer pointed to a different container. Across trials, the Knower always pointed to the baited container, and the Guesser always to one of the other three containers. The chimpanzee was allowed to search one of the containers and could keep the food if it were found. Two of the four animals tested rapidly learned to select the Knower’s container more often than the Guesser’s.

The *transfer stage* of the procedure was designed to determine whether the discrimination was established based on the trainers’ visual access to the baiting procedure. During this transfer stage, baiting was done by a third person in the presence of both the Knower and the Guesser. During the baiting procedure the Guesser wore a paper bag over his or her head. The consequences were the same if the chimpanzee located the food item. For each of the chimpanzees participating in the study, mean choice accuracy in the final 50 trials of stage 1 was comparable with that in the 30 trials of stage 2. However, at the beginning of the transfer stage performance was at chance level (Povinelli, 1994), suggesting that the animals learned a novel

discrimination between bagged and nonbagged trainers during the transfer stage rather than generalizing the discrimination acquired in stage one to the “untrained” task.

Heyes (1998) offered three recommendations for improving the methodology of the Guesser-Knower procedures. First, non-reinforced probe trials should be used during the transfer stage of the procedure to ensure that success on the transfer probes could not be attributed to the learning of a new discrimination as occurred in the original study. Second, both the Guesser and Knower could wear opaque and translucent goggles, respectively, rather than a bag-on-head manipulation for transfer trials. This should sharpen the discrimination being made based on the sight rather than the presence or absence of the paper bag. Third, a pre-training phase could be added in which the participants are exposed to the opaque and translucent goggles with distinctively colored rims. This phase is intended to pre-expose the participants to the relationship between wearing the opaque goggles and not being able to see, and the relationship between wearing the translucent goggles and being able to see. The latter two modifications would make it more likely that the animals could solve the problem using observable cues such as “eye-object line”, in other words, by choosing the trainer for whom there is an unobstructed line between their eyes and the baiting event.

Statement of the Problem

Given that 80% of children with ASDs fail standard FB tests and that speech deficits may be an obstacle in their ability to succeed on standard FB tests, research that examines both nonvocal FB tests and component perspective taking skills should generate useful information about perspective taking in children with ASDs. Therefore, the purpose of this research was to examine performance on: (a) standard FB tests, (b) nonvocal FB tests, and (c) the Guesser-Knower (GK) test with children diagnosed with ASDs.

I hypothesized that significantly more children would pass the nonvocal FB tests than the standard FB tests. Moreover, all children who passed the standard FB tests would also pass the nonvocal tests, but some or all children who failed the standard FB tests would pass the nonvocal FB tests. Further, I hypothesized that significantly more children would pass the GK test than either the standard FB or the nonvocal FB tests.

Method

Participants and Setting

Eighteen children, 16 boys and 2 girls, with diagnoses of ASDs were recruited from Winnipeg through St. Amant, Children Special Services, and the Manitoba Autism Research Team. Ethical approval was obtained from the Psychology/Sociology Research Ethics Board of the University of Manitoba and written informed consent to participate in the study was obtained from the legal guardians.

The mean chronological age of the children was 7 years 3 months (range = 5 years to 11 years 11 months). All children had prior confirmed diagnoses, which included 4 children with autistic disorder, 2 with PDD-NOS, and 4 with Asperger syndrome. The remaining 8 children had a general diagnosis of ASD. Diagnostic information was obtained from the parents for 15 children and from record review for the other three. Child characteristics are presented in Table 1.

To be included in the study, children were required to demonstrate a minimum verbal age of 36 months on expressive and receptive language based on their language assessment scores. Although previous research has suggested that typically developing children pass standard FB by age 4 (e.g., Baron-Cohen, 1985), Siegler (1986) reported that after age 3 children understand that even when another person's beliefs are false, people will behave according to their beliefs.

Table 1

Children's Characteristics and Test Results

Child	Age (yrs- mo)	Diagnosis	Sex	IQ	Expressive Language Standard Scores	Standard False Location	Standard False Content	Nonvocal False Location	Nonvocal False Content	Guesser Knower
1	5-03	Autistic Disorder	M	82	95	- ¹	-	-	-	-
2	11-11	Autistic Disorder	M	40	45	-	-	-	-	-
3	5-00	ASD	M	54	61	-	-	-	-	-
4	6-01	ASD	M	65	77	-	-	-	-	-
5	5-11	ASD	M	96	119	-	-	-	-	-
6	5-03	Autistic Disorder	M	80	65	-	-	-	-	-
7	5-06	ASD	M	76	55	-	-	-	-	-
8	6-07	ASD	M	73	61	-	-	-	-	-
9	7-05	PDD-NOS	F	81	61	-	-	-	-	-
10	6-11	Autistic Disorder	M	60	45	-	-	-	-	-
11	6-00	ASD	M	57	55	-	-	+	-	+
12	10-08	Asperger	F	94	98	-	-	+	+	+
13	9-05	ASD	M	59	65	+	-	-	-	-
14	8-05	ASD	M	100	89	+	+	+	+	-
15	6-05	Asperger	M	84	67	+	-	-	+	-
16	9-11	Asperger	M	87	108	+	+	+	+	+
17	9-05	Asperger	M	119	110	+	+	+	+	+
18	8-04	PDD-NOS	M	86	95	+	+	+	-	+

¹ - denotes a fail (incorrect responding on one or more of the reality, memory, and belief questions on the false belief tests or less than 8 correct out of 10 trials on the Guesser-Knower test); + denotes a pass (correct responding to all questions on the false belief tests or at least 8 correct out of 10 trials on the Guesser-Knower test).

Furthermore, Hala, Chandler, and Fritz (1991) reported that 80% of the 3-year-old children participating in their study correctly responded to a false belief prediction question on a hide and seek test. Children under 7 years of age were assessed using the Preschool Language Scale 4 (PLS-4; Zimmerman, Steiner, & Pond, 2002). The PLS-4 is designed to identify children who have a language disorder or delay and has been normed for children ages 0 months through six years 11 months. Children older than 7 years of age were assessed using the Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-4; Wiig & Secord, 2004). All but two of the 18 children (Children 3 and 11) had expressive language age equivalent scores above 48 months (mean = 78, range = 35 – 119).

Children under 7 years were assessed using the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III, Wechsler, 2002). The WPPSI-III is an individually administered clinical instrument for assessing the intelligence of children aged 2 years 6 months through 7 years 3 months. Children 7 years or older were assessed using the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV, Wechsler, 2004). IQ scores were not expected to correlate with FB performance given that children with Down syndrome who had lower IQ scores consistently outperformed children with ASDs who had higher IQ scores (Baron-Cohen et al., 1985; 1986). The IQ tests were conducted to allow for comparison with previous research. The mean IQ score across children was 77.39 (range = 40 to 119). IQ scores are presented in Table 1.

All assessments took place in a quiet assessment room at the St. Amant Research Centre or at a quiet location in the children's homes.

Apparatus

False location and false content standard FB tests. Materials for the Bobby Anne False

Location FB test included two puppets (Bobby and Anne), a styrofoam ball, a wooden basket with a lid, and a wooden heart shaped box with a lid. Materials for the False Content FB test included an empty Smarties box and a pencil.

False location and false content nonvocal FB tests. The materials for the nonvocal FB tests were similar to those used in the two standard FB tests described above with the following modifications and additions. The nonvocal FB False Location test included a small green frog, a purple plastic cylindrical container with a lid, a yellow purse with a zipper, and two 7.5 x 12.7 cm index cards. One card was green and had the word "YES" printed on it. The other was red and had the word "NO" printed on it. The nonvocal FB False Content test included an empty crayon box, a playing card, a crayon and the two Yes/No index cards.

Guesser-Knower test. The materials for the GK test included: two wooden boxes with hinged covers identical in size, shape, and colour; a 25 cm high x 1 m wide partition; a red wooden star; and two pairs of goggles, one pair with orange rims, and the other with purple rims. One pair of goggles was opaque, the other translucent. The orange goggles were opaque for half of the children and translucent for the other children.

Experimental Design

All children were tested once on the two standard FB tests (False Location and False Content); and the two corresponding nonvocal FB tests. The order of testing was counterbalanced to control for order effects. Children were randomly assigned to one of the following eight sequences:

- (1) Standard Bobby Anne, Standard Smarties, Nonvocal Bobby Anne, Nonvocal Crayon;
- (2) Standard Bobby Anne, Standard Smarties, Nonvocal Crayon, Nonvocal Bobby Anne;
- (3) Standard Smarties, Standard Bobby Anne, Nonvocal Bobby Anne, Nonvocal Crayon;

- (4) Standard Smarties, Standard Bobby Anne, Nonvocal Crayon, Nonvocal Bobby Anne;
- (5) Nonvocal Bobby Anne, Nonvocal Crayon, Standard Bobby Anne, Standard Smarties;
- (6) Nonvocal Bobby Anne, Nonvocal Crayon, Standard Smarties, Standard Bobby Anne;
- (7) Nonvocal Crayon, Nonvocal Bobby Anne, Standard Bobby Anne, Standard Smarties; and
- (8) Nonvocal Crayon, Nonvocal Bobby Anne, Standard Smarties, Standard Bobby Anne.

Moreover, half the children were tested on the GK test after the standard FB and nonvocal FB tests, and the other half were tested first on the GK test, followed by the standard FB and nonvocal FB tests. Depending on their attending behaviours, some children's assessments were conducted over two home visits while others completed all assessments in one home visit of approximately 1.5 hours.

Standard False Belief Assessment Procedures

All children were tested on two standard FB tests (Bobby Anne False Location test and the Smarties False Content) using the procedures described by Baron-Cohen et al. (1985) and Perner et al. (1989), respectively. The child was seated directly across the table from the experimenter who puppeteered the dolls and carried out the procedures following the scripts shown in Appendix A. Following child responses the subsequent question was delivered, and following the final question the test was completed and the child was told that he or she was done and was thanked for helping. On the Bobby Anne False Location test a pass was scored if the child responded correctly on the *belief* question ("Where will Bobby look for the marble?"), the *reality* question ("Where is the marble really?"), and the *memory* question ("Where was the marble in the beginning?") according to Baron-Cohen et al. If the child responded incorrectly to any of these questions a fail was scored. If the child did not respond within 15 s, the question was repeated. If the child did not respond a second time, the question was scored as incorrect and

the experimenter proceeded to the next question. Nonvocal responses (such as pointing to the container and head nodding in response to questions that required yes/no answers) were recorded and accepted.

On the Smarties test, the child was seated directly across the table from the experimenter, who carried out the procedures as described in Appendix A. A pass was scored if the child responded correctly on the *prediction* question ("What will [name] say is in the box?"), the *reality* question ("What is really in the box?"), and the *memory* question ("Do you remember, when I took the box out of my bag and asked you what was in it, what did you say?"). Nonvocal responses (such as head nodding or shaking) to questions that required yes/no answers (e.g., "Is Jon here with us?") were accepted.

Nonvocal False Belief Assessment Procedures

Two nonvocal FB tests were administered. The nonvocal FB tests were similar to the standard FB tests, but the child was instructed to respond nonvocally to all questions. The detailed scripts of the nonvocal FB testing procedures are shown in Appendix B. For the prompt questions that required a yes or no response, the child was presented with two cards with the word "YES" printed on one and the word "NO" printed on the other. The child was asked to communicate their responses by pointing to or touching one of the cards. Prior to conducting the nonvocal FB tests all children were tested using the yes/no cards. In order to begin the nonvocal FB test the child had to respond correctly to five questions for which the answer was yes and five questions for which the answer was no (e.g., showing the child a spider and asking "is this a spider?" or showing the child a fork and asking "is this a spoon?"). All but one child (i.e., child 11) completed the testing using the Yes/No cards. Child 11 received 14 training trials, 12 of which required prompting to use the cards, and two of which were independent. Because he was

responding correctly vocally (i.e., yes or no) and only required a prompt to “use the cards” we proceeded with the assessment. On average children received 14 training trials (range = 10-42) before the nonvocal FB assessments were conducted. During the pre-testing with the yes/no cards all of the children discriminated between the two cards and were able to respond correctly to questions requiring yes or no answers using the cards. For the belief question on the False Location test, the child was presented with the purple container and the yellow purse. For the belief question on the False Content test, the child was presented with a crayon and a playing card. For both tests the child was required to communicate their responses by pointing to or touching the container of their choice. However, if vocal responses were emitted, they were recorded and accepted.

Guesser-Knower Assessment Procedures

Each child was exposed to two phases: goggle preexposure and assessment. Prior to each session the child was presented with a tray of edibles from which they were instructed to select one. The edibles used were items identified as preferred by the child's parents. The item chosen by the child was used to reinforce attending on each trial independent of the child's responding. For the purpose of this study attending was defined as looking at me while remaining in their seat.

Goggles preexposure. Before the preexposure phase began, children were presented with a number of items and asked to label those items. I held up each item in front of the child and said “what's this?” Once the child correctly labelled the items to be used, the goggle preexposure phase began. During this phase a child was exposed to the two pairs of goggles. At the beginning of each trial, the selected edible was given to the child for looking at me while remaining in the seat. Once the edible had been consumed, the child was given the goggles and asked to put them

on, with my help when necessary. The child was only required to wear the goggles briefly (approximately 1-2 s) and was able to remove the goggles at any time. However, the child was allowed to keep the goggles for the duration of the trial. While wearing the goggles or following their removal, the child was asked to point to or interact with an object that I had placed on the table (e.g., touch it, pick it up, or hand it to me). Selected items were those that the child was able to correctly label. The child was given up to 15 s to respond. After the child had attempted or completed the response, or after 15 s without a response, I thanked the child for helping (regardless of whether the child had responded successfully), waited approximately 15 s, and presented the next trial. All but three children completed a minimum of six preexposure trials (mean number of trials was 10).

Children 2 and 11 refused to wear both pair of goggles and child 7 would only wear the translucent goggles. Therefore, I put the goggles on and asked the children whether they could see my eyes. Once the children responded correctly by saying "no" when I was wearing the opaque goggles and "yes" when I was wearing the translucent goggles, the assessment began.

Guesser-Knower assessment. During the assessment two assistants sat across the table facing the child, and I sat to the left or right of the child at the end of the table. On the table were two identical boxes in front of the child with a plexiglass barrier placed over the boxes. Pre-trial reinforcement for attending was conducted as in the goggle preexposure phase. Following reinforcement for attending, a partition was placed between the child and the boxes, such that the child was unable to see the boxes. The upper bodies of the assistants remained visible. In view of the child, one assistant put on the translucent goggles (the Knower) and the other put on the opaque goggles (the Guesser) and both assistants oriented their heads towards the boxes. I held up the wooden star to show it to the assistants and the child and said, "I'm going to hide the star."

I then opened and closed each box in sequence behind the partition and went through the motion of placing the wooden star in each box, although the star was actually placed in only one of the boxes. The box with the star varied across trials according to a predetermined randomized order. The partition was lowered and each assistant pointed to a different box. The Knower always pointed to the box with the star and the Guesser to the box without the star. The child was asked, "Who knows where the star is?" and was given up to 15 s to respond. A correct response was defined as touching or pointing to the Knower or the corresponding box. An incorrect response was defined as touching or pointing to the Guesser or the corresponding box. Once the child responded, I opened the corresponding box and showed the contents to the child. I thanked the child for helping regardless of the accuracy of the child's response, waited approximately 15 s, and then presented the pre-trial reinforcement and the next trial. If the child did not respond within 15 s I repeated the question "Who knows where the star is?" and again, the child was given up to 15 s to respond. If the child did not respond the second time, the trial was terminated with no programmed consequences and an incorrect response was recorded. Each test session consisted of 10 trials. The role of the assistants as Guesser and Knower were counterbalanced across trials. A pass was awarded on the GK test if the child responded correctly on a minimum of 8 out of 10 trials.

Interobserver Reliability and Procedural Integrity Assessments

Interobserver reliability. An observer independently recorded the child's responses on each FB test question and on each GK test trial. My recordings on each child's response were compared with the observer's. During the FB tests, an agreement occurred if both the observer and I agreed that the child passed the test by responding correctly on the Belief, Reality, and Memory questions. A disagreement occurred if we disagreed on any of the three questions.

During the GK test, an agreement was scored if the observer and I recorded the same response by the child on each trial. Otherwise, a disagreement occurred. Interobserver agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (Kazdin, 1982; Martin & Pear, 2007).

Interobserver reliability checks were conducted for 72% of the Standard False Location tests, 67% of the Standard False Content tests, 83% of the nonvocal False Location tests, 61% of the nonvocal False Content tests, and 38% of the GK trials across 38% of the children. Interobserver agreement scores were 100% for the Standard False Location, 83% for the Standard False Content, 93% for the nonvocal False Location, 100% for the nonvocal False Content, and 98% (range = 90%-100%) for the GK tests. Of the three disagreements that occurred across FB tests, all children failed the respective tests and none of the disagreements would have changed the children's performances.

Procedural integrity. An observer occasionally recorded whether I correctly administered the FB tests. These checks were made on 61% of Standard False Location tests, 61% of Standard False Content tests, 67% of nonvocal False Location tests, and 50% of nonvocal False Content tests. An FB test was scored as correctly administered if I had followed every step described in the scripts (Appendices A and B) and incorrectly administered if I had deviated from the procedures. The percentage of FB tests correctly administered were 91% for the Standard False Location, 100% for the Standard False Content, 100% for the nonvocal False Location, and 89% for the nonvocal False Content tests.

For the GK test, an observer scored whether I delivered each trial correctly using a procedural checklist (Appendix C). These checks were conducted for 38% of the children (10 trials per child). A trial was scored as correctly administered if I had followed all steps on the

checklist; otherwise, it was scored as incorrect. The average percentage of trials delivered correctly per child was 90%, ranging from 70% to 100%.

Results

The results of the 18 children are shown in Table 1. Four of the 18 children (22%) passed both standard FB tests, 12 of the 18 children (67%) failed both tests, and 2 of the 18 children (11%) passed the False Location but failed the False Content test. Four of the 18 children (22%) passed both nonvocal FB tests, 11 failed both (61%), and 3 passed one of the tests (17%). Lastly, 5 of the 18 children (28%) passed the GK test.

Of the 4 children who passed both standard FB tests, 3 passed all nonvocal FB tests while the fourth child passed the nonvocal False Location test but failed the nonvocal False Content test. Of the 12 children who failed both standard FB tests, 1 child passed both nonvocal FB tests, and 1 child passed one nonvocal FB test. Of the two children who passed only one standard FB test one went on to pass one of the nonvocal FB tests and the other failed both. All other children failed the nonvocal FB tests.

Three of the 4 children who passed both standard FB tests passed the GK test, obtaining scores between 90% and 100%. The fourth child (i.e., child 14) obtained a score of 40% on the GK test.

Of the 12 children who failed both standard FB tests, 2 passed the GK, obtaining scores of 80% and 100%, respectively. The other 10 children had scores ranging from 20% to 70% with the majority in the 40%-60% range. Children 13 and 15, who each passed one standard FB test, scored 50% on the GK test.

Performance on the standard FB and nonvocal FB tests by question is presented in Table 2. On average 90% (range = 89%-94%) of the children responded correctly to the reality

questions, 51% (range = 39%-61%) responded correctly to the memory questions, and 42% (range = 39%-50%) responded correctly to the belief questions. Percent correct on the belief and reality questions were similar across all FB tests, while there was slightly more variability in responding on the memory questions across the different tests. Performance on the standard FB tests replicate previous research (i.e., 67%-78% of children with ASDs failed the standard FB tests). Similar results were found for the nonvocal versions of the test. Twenty-eight percent of the children passed the GK test.

Table 2

Percent correct responses to questions on false belief tests and percent of children who passed the test.

	Belief	Reality	Memory	Percent Pass
Bobby Anne False Belief	39%	94%	61%	33%
Smarties False Belief	39%	89%	39%	22%
Nonvocal Bobby Anne False Belief	39%	89%	61%	33%
Nonvocal Smarties False Belief	50%	89%	44%	28%

For statistical analysis, the results were collapsed such that children who passed one or both of the standard FB tests were assigned a pass and those who failed both were assigned a fail. The same was done for the nonvocal FB tests. To test whether performance differed across the FB and GK tasks, Cochran's Q test (Morgan, Glinger, & Harmon, 2003) was used to compare the three assessments. There was no significant difference on the percentage of children who passed each test ($Q(2) = 1.200, p = .549$). However, using a point biserial correlation, expressive language standard scores were correlated positively with performance on the standard FB, nonvocal FB, and GK tests, although only the latter was statistically significant (standard FB, $r_{pb} = .40, p = .10$; nonvocal FB, $r_{pb} = .45, p = .06$; GK, $r_{pb} = .47, p = .05$).

In a post-hoc analysis of nine children (1, 3, 4, 6, 10, 13, and 15 through 17) whose nonvocal FB assessments were videotaped, 36% of responses were vocal, 34% of responses were nonvocal, and 30% of responses were a combination of vocal and nonvocal. Among the 30% of vocal and nonvocal responses, there was agreement between the two modes of responding.

The nine children was comprised of two who passed both nonvocal FB tests (children 16 and 17), one who passed only the nonvocal False Content test (child 15), and six who failed both tests. Child 16 responded to 7% of the questions vocally, 57% nonvocally, and 36% both vocally and nonvocally. Child 17 responded to 64% vocally, 29% nonvocally, and 7% using a combination of vocal and nonvocal responses. Child 15, who passed only the nonvocal False Content test responded to 53% of the questions vocally and 47% nonvocally. The six children who failed both tests responded to 32% of the questions vocally, 28% nonvocally, and 40% using a combination of both vocal and nonvocal responses.

A post-hoc analysis was conducted to examine the relationship between the working memory subtest on the WISC-IV and performance on the FB tests. Twelve children (2, 4, 8, 9, 10, and 12 through 18) had the WISC-IV administered (the WPPSI-III does not produce a working memory index score). Seven of the 12 children had passed at least one of the FB tests and five of those seven children's working memory index standard scores fell within one standard deviation of the mean (i.e., between 85 and 115), one had a standard score of 121 and the other a standard score of 80. Among the five children who failed all FB tests, two obtained standard scores that were between two and three standard deviations below the mean (i.e., between 55 and 70), two obtained scores that were below three standard deviations below the mean (i.e., below 55) and one scored a 91.

In an additional examination of performances on the five tests a pattern emerged among

the four children whose confirmed diagnoses were autistic disorder and the four whose diagnoses were Asperger syndrome. All four children who had a diagnosis of autistic disorder failed all standard FB, nonvocal FB and GK tests. Alternatively, of the four children who had a diagnosis of Asperger three passed at least one of the standard FB tests, all of them passed at least one of the nonvocal FB tests, and three also passed the GK test. The performance of children with autistic disorder showed that they performed well on the reality question but poorly on the belief and memory questions.

Discussion

I hypothesized that significantly more children would pass the nonvocal FB tests than the standard FB tests because the former did not require expressive vocal language. This hypothesis was not supported. However, the pass/fail patterns showed that two children passed the nonvocal FB tests and failed the standard FB tests, and five of the six children who passed an standard FB test also passed an nonvocal FB test. Overall, these results suggest that the nonvocal FB tests may only be slightly easier than the standard FB tests at best, and that using nonvocal instead of vocal responses did not significantly improve test performance for this sample.

I also hypothesized that significantly more children would pass the GK test than the FB tests. This hypothesis was not supported. Performance on the GK test did not differ significantly from performance on the standard FB or nonvocal FB tests. Therefore, it remains unclear how "seeing and knowing", as measured by the GK test, is related to false belief task performance.

Several observations from the results of this study are consistent with previous research. Firstly, 22% of the current sample passed both standard FB tests (Baron-Cohen et al., 1985, 1986; Perner et al., 1989; Swettenham, 1996). Approximately 90% of all children who participated in this study were able to respond correctly to the reality question on both FB tests,

while approximately 40%-50% responded correctly to the belief question (Table 2). This finding is consistent with previous research (e.g., Baron-Cohen et al. and Perner et al.) which found that all children passed the reality question while only 17%-20% of the children correctly responded to the belief question. Approximately 40%-60% of the children passed the memory question on both tests in this study, a finding which is similar to that of Perner et al. who reported 61% of children passing the memory question among their sample. This contradicts results presented by Baron-Cohen et al., who reported that 100% of children responded correctly to the memory question on the Sally Anne test regardless of performance on the belief question.

Among the children who failed the standard FB and nonvocal FB tests, more children responded correctly to the memory question on the False Location tests than on the False Content tests. It is possible that the children showed better performance on the False Location memory question because responses were prompted by visual cues from the task materials, whereas children were asked to recall what they had said on the False Content test. Based on this observation in conjunction with the results of the post-hoc analysis of working memory and FB test performance, future research might further investigate this relationship by restricting inclusion criteria to children whose working memory, as measured by standard intellectual assessments (e.g., WISC-IV), does not fall more than one standard deviation below the mean. Alternatively, children with higher and lower scores on the WISC-IV working memory subtest could be compared while controlling for expressive language scores.

Another finding that is consistent with previous research was the positive moderate correlations between expressive language standard scores and performance on the standard FB and nonvocal FB tests. This finding is consistent with previous research which looked at verbal mental age (VMA) and performance on standard FB tests. Happé (1995) conducted an analysis

on data collected between 1985 and 1993 which looked at the relationship between VMA (including both expressive and receptive language) and performance on perspective taking tests. Results of that analysis showed that VMA was a good predictor of, and was highly correlated with, performance on perspective taking tests among both children with ASDs and typically developing children. Although the sample in the present study is small, the results support Happé's finding of a positive correlation between language scores standard FB test performance.

The observed differences between performance on the standard FB, nonvocal FB, and GK tests by children whose diagnosis was autistic disorder versus those whose diagnosis was Asperger syndrome highlights a need for further examination of this variable. Previous research has reported that individuals diagnosed as having Asperger syndrome perform better on perspective taking tests than other individuals with autistic disorder diagnoses (e.g., Bowler, 1992). One possibility is that children with Asperger syndrome have higher language scores.

The difference between performance on the standard FB and nonvocal FB tests was not significant. Furthermore, providing nonvocal response alternatives on the nonvocal FB test did not improve the children's performance in this study. Although children were asked to respond using nonvocal methods of communication during the nonvocal FB tests, the results of nine children's responses during the nonvocal assessments indicated that all of the children responded to at least some questions vocally. It is possible that the language skills of the children in the study were too well developed and therefore they did not benefit from the addition of nonvocal response options. Perhaps future studies could examine nonvocal response options with children with lower verbal age scores.

Finally, I found no performance difference between the GK test and the FB tests. The GK test used during the present study improved upon previous research in that the seeing and

knowing discrimination was refined from an individual wearing a paper bag over their head to individuals wearing opaque and translucent goggles. This had the advantage of limiting the discrimination to the Guesser's and Knower's eyes. However, it is possible that children could still pass the test based on the visual discrimination between the purple and orange goggle rims associated with the roles of the Guesser and Knower throughout the test. Perhaps future versions of the test might consider alternative methods for enhancing child attendance to the eyes of the Guesser and Knower while minimizing superfluous stimuli. These test variations might include hand covered eyes versus uncovered eyes, closed versus opened eyes, the Guesser turning around and the Knower facing the child, etc. Furthermore, future researchers might conduct multiple versions of the test to ensure that the children who pass are making the relevant discrimination between seeing and not seeing.

In summary, it appears that language ability and memory strongly influenced the performance of children with ASD on the FB tasks. Nonvocal response options were not useful for children with verbal age equivalents of 4 years. Lastly, further refinement of the Guesser-Knower task is needed.

References

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, text revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association.
- Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders, 18*, 379-402.
- 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., Leslie, A. M. & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113-125.
- 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* (5), 407-418.
- Baron-Cohen, S., Wheelright, 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.
- Bowler, D. M. (1992). "Theory of Mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 33*, 877-893.
- Charlop-Christy, M. H., & Daneshvar, S. (2003). Using video modeling to teach perspective

- taking to children with autism. *Journal of Positive Behavior Interventions*, 5 (1), 12-21.
- Dahlgren, S., & Trillingsgaard, A. (1996). Theory of mind in non-retarded children with autism and AS. A research note. *Journal of Child Psychology and Psychiatry*, 37, 759-763.
- Fein, D., Pennington, B., Markowitz, P., Braverman, M. & Waterhouse, L. (1986). Towards a neuropsychological model of infantile autism: Are the social deficits primary? *Journal of the American Academy of Clinical Psychiatry*, 25, 198-212.
- Flavell, J., Botkin, P. T., Fry, C. L., Wright, J. W. & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.
- Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders*, 33(4), 365-382.
- Gopnick, A., & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development*, 62, 98-110.
- Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development*, 62, 83-97.
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6(3), 346-359.
- 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. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843-855.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21, 101-148.

- Hogrefe, G. J., Wimmer, H., & Perner, J (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 57, 567-582.
- Holroyd, S., & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders*, 23, 379-385.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Leslie, A. M., Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, 6, 315-324.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, 74 (4), 1130-1144.
- Martin, G. L., & Pear, J. J. (2007). *Behavior modification: What it is and how to do it* (8th ed.) Upper Saddle River, NJ: Prentice Hall.
- Morgan, G. A., Glinger, J. A., & Harmon, R. J. (2003). Selection and use of inferential statistics: A summary. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42 (10), 1253-1257.
- Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Borselenberg, R., van den Hogen, T., et al. (1999). The TOM test: A new instrument for assessing theory of mind in normal children and children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 29 (19), 67-80.
- Olson, D. R. (1988). On the origins of beliefs and other intentional states in children. In J. W. Astington, P. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 414-426). Cambridge, MA: Cambridge University Press.

- Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology*, 6, 415-431.
- Pelios, L. V., & Lund, S. K. (2001). A selective overview of issues on classification, causation, and early intensive behavioral intervention for autism. *Behavior Modification*, 25: 678-697).
- Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development*, 60, 689-700.
- Povinelli, D. J. (1994). Comparative studies of animal mental state attribution: A reply to Heyes. *Animal Behaviour*, 48, 239-241.
- Povinelli, D. J., Nelson, K. E. & Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 104, 203-210.
- Rutter, M. (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry*, 24, 513-531.
- Siegler, R. S. (1986). *Children's thinking*. Upper Saddle River, NJ: Simon and Schuster.
- Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry*, 37, 157-165.
- Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence 3rd ed.* Texas: The Psychological Corporation.
- Wechsler, D. (2004). *Wechsler intelligence scale for children 4th ed.* Texas: The Psychological Corporation.
- Wiig, E. H., & Secord, W. A. (2004). *Clinical evaluation of language fundamentals 4th ed.* Texas: The Psychological Corporation.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale 4th ed.*

Texas: The Psychological Corporation.

Appendix A

Scripts for Standard False Belief tests

“Bobby-Anne” False Location False Belief test

The trainer will place a box, a basket, and the two dolls in front of the child. The trainer then will tell the following story, moving the dolls accordingly:

The experimenter will point to each of the dolls and name them. This is Bobby and Anne.

Naming question: Who are these?

Correct Answer: Bobby and Anne

Bobby and Anne are good friends. One day they found a ball and decided to keep it in the basket. Experimenter guides Bobby through motions to place the ball in the basket in view of the child.

Prompt question 1: Where did Bobby and Anne keep the ball?

Correct Answer: Basket.

Bobby says, “Excuse me. I need to go to the washroom.” So off he goes. A second experimenter will take Bobby out of the room.

Anne takes the ball from the basket and hides it in the box.

Prompt question 2: Is Bobby here?

Correct Answer: No.

Prompt question 3: Did Bobby see Anne hide the ball in the box?

Correct Answer: No.

Prompt question 4: Does Bobby know what Anne did?

Correct Answer: No.

Bobby comes back. He says, “I am going to get the ball.”

Belief question: Where will Bobby look for the ball?

Correct Answer (1 point): Basket.

Reality question: Where is the ball really?

Correct Answer (1 point): Box.

Memory question: Where did Bobby and Anne put the ball before Bobby left?

Correct Answer (1 point): Basket.

“Smarties” False Content False Belief test

After introducing an assistant (Jon) to the child, Jon will leave the room. The trainer will hold up a box of “Smarties” and say the following: “Look, I have a box of something here.”

Prompt question 1: What do you think is inside this box?

Correct answer: Smarties, candy, chocolate, or sweets.

Let’s take a look. Oh! It’s a pencil!

Reality question: What is really inside this box?

Correct answer (1 point): Pencil.

Prompt question 2: Is (Jon) here with us?

Correct answer: No.

Prompt question 3: Does (Jon) know that there is a pencil inside this box?

Correct answer: No.

Here comes (Jon)!

Belief question: What does (Jon) think is inside the box?

Correct answer (1 point): Smarties, candy, chocolate, or sweets.

Memory question: “Do you remember when I took the box out of my bag and asked you what was in it, what did you say?”

Correct answer: What the child said in Prompt Question 1.

Appendix B

Scripts for Nonvocal False Belief tests

“Bobby-Anne” Nonvocal False Location False Belief test

The trainer will place the two containers and the two dolls in front of the child. The trainer then will tell the following story, moving the dolls accordingly:

The experimenter will point to each of the dolls and name them. “This is Bobby and Anne.”

Naming question: The experimenter will ask the child to “Point to Bobby” and “Point to Anne.”

“Bobby and Anne kept this frog in this purple container.” The frog is placed in the container in view of the child.

The experimenter will present the child with the purple container and a yellow purse and ask the prompt question.

Prompt question 1: Where did Bobby and Anne keep the frog?

Correct response. (1 point) The child will point to or touch the purple container.

Anne says, “Excuse me. I need to go to the washroom.” So off she goes. A second experimenter will take Anne out of the room.

Bobby takes the frog out from the purple container and hides it in the yellow purse.

The experimenter will place a green card with the word YES printed on it and a red card with the word NO printed on it on the table, and then ask the prompt question.

Prompt question 2: Is Anne here?

Correct answer: Child will point to or touch the NO card.

The experimenter will place a green card with the word YES printed on it and a red card with the word NO printed on it on the table, and ask the next question.

Prompt question 3: Did Anne see Bobby move the frog to the purse?

Correct answer: Child will point to or touch the NO card.

The experimenter will place a green card with the word YES printed on it and a red card with the word NO printed on it on the table, and ask the next question.

Prompt question 4: Does Anne know what Bobby did?

Correct answer: Child will point to or touch the NO card.

Anne comes back. She says, “I am going to get the frog.”

Experimenter will present the child with the purple container and the yellow purse and ask the belief question.

Belief question: Where will Anne look for the frog?

Correct response (1 point): The child will point to or touch the purple container.

Reality question: Where is the frog really?

Correct Answer (1 point): The child will point to or touch the yellow purse.

Memory question: Where did Bobby and Anne put the frog before Anne left?

Correct Answer (1 point): The child will point to or touch the purple container.

“Crayon” Nonvocal False Content False Belief test

After introducing an assistant (Jon) to the child, Jon will leave the room. The trainer will hold up a box of “Crayons” and say the following: “Look, I have a box of something here.”

Prompt question 1: What is in this box?

The experimenter will place a crayon and a playing card on the table.

Correct answer: The child will point to or touch the crayon.

Let’s take a look. Oh! It’s a card!

Reality question: What is in this box?

The experimenter will place a crayon and a playing card on the table.

Correct answer (1 point): The child will point to or touch the playing card.

Prompt question 3: Is (Jon) here?

The experimenter will place a green card with the word YES printed on it and a red card with the word NO printed on it on the table.

Correct answer: Child will point to or touch the NO card.

Prompt question 4: Will (Jon) say that there is a card in the box?

The experimenter will place a green card with the word YES printed on it and a red card with the word NO printed on it on the table.

Correct answer: Child will point to the NO card.

“Here comes (Jon)!”

Belief question: What will (Jon) say is in the box?

The experimenter will place a crayon and a playing card on the table.

Correct answer (1 point): The child will point to or touch the crayon.

Memory question: Experimenter asks “When I first took the box out of my bag, what did you say was in it?”

The experimenter will place a crayon and a playing card on the table.

Correct answer (1 point): Same response to Prompt Question 1.

Appendix C

Procedural Checklist for Guesser Knower Test

1. Experimenter gave edible for attending behaviour
2. Partition set up
3. Guesser and Knower put on correct goggles
4. Experimenter holds up star and says "I'm going to hide the star"
5. Experimenter opens and closes each box in order
6. Experimenter places the star in the correct box
7. Partition is lifted
8. Guesser and Knower point to the correct boxes
9. Experimenter asks the child "Who knows where the star is?"
10. Child is given 15 s to respond. If no response trial is terminated after 15 s.
11. Experimenter opens corresponding container and shows the contents
12. Experimenter remains neutral following child response
13. Experimenter thanked the child for helping
14. Experimenter waits 15 s before starting a new trial.