Social Cognitive Development During the Latency Years:
An examination of three measures

Matthew W. Mosconi

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology (Clinical Psychology).

Chapel Hill
2004

Approved By:

Advisor: Gary Mesibov, Ph.D.

Reader: Marilyn Hartman, Ph.D.

Reader: Steven J. Reznick, Ph.D.

Abstract

Matthew W. Mosconi

Social Cognitive Development During the Latency Years:
An examination of three measures

(Under the direction of Gary Mesibov, Ph.D.)

A host of research has focused on social cognitive development. Still, measurement limitations have hindered progress. Limitations include failures to examine developments in latency aged children and a scarcity of data examining related cognitive skills. We administered three social cognition measures, the Faux Pas Task (FP), the Movie Stills task (MS), and the Animations task, to 64 children between 6-12 years, and assessed children's verbal ability. Results indicated that chronological age was related to MS and Animations task performance, but not to FP task performance after covarying children's verbal ability. FP task performance was related to verbal mental age. Individual items of the FP task were separated in difficulty by whether or not they contained a negative word or phrase. Items with a negative word or phrase were answered correctly more often. These results and their implications for future research on social cognitive development are discussed.

To Mom, Dad, Chris and Laura. Whenever I do good, it's because of you.

Acknowledgments

I owe a great deal of thanks to Gary Mesibov for his constant support, clear guidance, and ability to push without applying pressure. I can't imagine a more comforting mentor through the graduate experience, and I thank you for that.

I also owe Steve Reznick my gratitude. Your ability to think in a novel and critical manner is something that I will carry with me throughout my professional career. More importantly, your kind support and encouragement are appreciated and impressive.

Finally, thank you Marilyn Hartman for your support and for providing a model of diligence and attention to detail.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Background and Significance

A great deal of research has explored social cognition (i.e., the ability to interpret social information). This research has highlighted the importance of studying social cognitive development for understanding changes in social behavior and developments in language and other cognitive abilities. Additionally, research has highlighted a connection between social cognitive deficits and several psychiatric disorders, such as autism (e.g., Baron-Cohen, 1995) and conduct disorder (Dodge & Pettit, 2003). In spite of the importance of these processes to cognitive and behavioral development, the measurement of social cognition has been studied infrequently. As a result, measurement limitations have hindered progress in the investigation of social cognitive development and, also, have drawn criticism from numerous researchers (e.g., Ozonoff, personal communication). Consequently, the goal of the present investigation was to examine three central social cognition tasks to determine their sensitivity to important social cognitive developments that occur in latency-aged typically developing children.

Social Cognitive Development

Social cognition refers to the ability to organize and interpret social information in order to understand and predict the behavior of others. It is characterized by a variety of skills, including the ability to infer the mental states of others, understand communicative gestures (e.g. nonverbal gestures), detect others' intentions, and recognize and regulate emotions (for a review, see Flavell, 1999). This paper will examine tasks measuring several of these functions, including theory of mind processing and emotion recognition.

Numerous investigations have explored the development of social cognitive skills. For example, researchers have shown that early in ontogeny, infants evidence a number of social cognitive precursors, including a preference for social stimuli relative to objects (Morton & Johnson, 1991), increased attention to faces relative to other stimuli (IBID), and the ability to discriminate different emotions (Bornstein & Arterberry, 2003). During the toddler and preschool years, children begin to display new social cognitive skills, including joint attention and theory of mind processing (i.e., the ability to infer the mental state of others). Then, throughout the latency years, they show even more complex social cognitive aptitudes, including the ability to integrate multiple pieces of information from intricate social interactions, decipher the intentions underlying social behavior, and recognize emotions from a broad set of dynamic and static cues. The present investigation will focus on measuring several of the social cognitive skills that develop during the latency years, including theory of mind processing and emotion recognition.

Advanced Theory of Mind Functions

Theory of mind processing refers to the ability to infer the mental states of others, including their thoughts, desires, beliefs, and intentions (Baron-Cohen et al., 2000). In young children, the emergence of a theory of mind is often triggered by the realization that others may have separate beliefs or a different perspective than their own. A lack of theory of mind is often demonstrated by young children (i.e., younger than age 4 years) who may wave good-bye with their hands facing themselves, replicating the perspective that they have when others are waving to them. Subsequent, or advanced theory of mind developments include understanding the intentions underlying behavior, and the comprehension that information not only is perceived, but also interpreted by others. For instance, Wellman and Hickling

(1994) suggested that four to five year old children only begin to understand that mental contents are actively constructed by the person and “are subject to biases, misrepresentations, and active interpretation” (p. 1578).

Research on advanced theory of mind functioning has been limited. Carpendale and Chandler (1996) suggest that advanced theory of mind processing largely has been ignored by researchers because of the mistaken assumption that children’s ability to identify that others have separate mental states is reflective of an understanding that people may interpret the same stimulus in different ways. This “interpretive” component of a theory of mind allows for children to appreciate the fact that one stimulus event can support two or more distinct interpretations. Latency-aged children improve in their ability to understand ambiguous behavior or events. Throughout development, children also become better able to understand more complex social environments. Complexity has been defined in terms of the number of variables that are related in a cognitive representation (Halford, Wilson, & Phillips, 1998) and has been shown to modulate the age at which children are able to pass theory of mind tasks (Andrews, Halford, Bunch, Bowden, & Jones, 2003).

Theory of mind development also is related to advancements in other cognitive domains. Research has indicated that success on theory of mind tasks may be mediated by general cognitive improvements. For instance, several studies have indicated that performance IQ is highly correlated with processing complex visual scenes (De Sonneville, et al., 2002) while verbal IQ (VIQ) may be predictive of performance on many verbally administered theory of mind tasks (Happe, 1994). As a result, developments in theory of mind often have been difficult to isolate. Thus, studies of advanced theory of mind development must adequately examine the development of the interpretive character of knowledge, the role of the complexity of information that has to be integrated in theory of mind tasks, and the influence of general cognitive development on performance.

Emotion Recognition Functions

In addition to theory of mind processing, several other social cognitive functions continue to develop throughout childhood. Amongst these functions is the ability to recognize others' emotions. Individuals use a variety of cues to infer emotions, including voice intonations, context, facial cues and nonverbal gestures. Emotion recognition skills appear to develop parallel to traditional theory of mind abilities (Buitelaar & van der Wees, 1997; Joseph, 1992) and are critical to later social aptitude (Iannotti, 1985) and prosocial behavior, such as empathy (Strayer, 1980).

Studies examining the development of emotion recognition in typically developing children have suggested that the processing of facial cues and other nonverbal emotional expressions undergoes major developments between 3-12 years old (Harrigan, 1984; Phillipot & Feldman, 1990). Several investigations have examined children's ability to correctly identify the emotions of characters in a story and have indicated that children improve in their ability to recognize emotions from facial and other nonverbal cues during latency years, and that females demonstrate superior performance relative to age-matched males (Buitelaar & van der Wees, 1997; Odom & Lemond, 1972; Rotenberg & Sullivan, 2003). Studying the development of emotion recognition skills during this period of childhood is thus important for both understanding important gender differences in social development and mapping typical developmental pathways of social processing.

Measuring Social Cognitive Development

A variety of measurement tools have been developed to assess aspects of social cognition. Most studies of theory of mind processing have incorporated false-belief tasks that were designed to assess individuals' ability to understand that people may have separate mental states. The following is a typical example of a false-belief task:

> *Sally and Ann are standing next to one another. Sally has a marble in the basket in front of her. Sally leaves the room and, while Sally is out of the room, Ann removes the marble and hides it in a different basket. Children are then asked, "When Sally returns, where will she look for the marble?"*

By four to five years of age, most children are able to accurately predict that Sally will look for the marble where she left it. This response signifies that children have an understanding that other individuals may have separate beliefs or knowledge states than their own. Thus, although the child in the experiment knows where the marble is, the child with a theory of mind also knows that 'Sally' does not know where the marble is.

Other social cognition measures have focused on children's interpretive character of knowledge (Carpendale and Chandler, 1996; Taylor, Lussier and Maring, 2003), understanding of humor (e.g., McGhee, 1979), sense of irony (e.g., Winner & Leekam, 1991), emotion recognition (Leppanen and Hietanen, 2001), and identification of socially awkward moments (Baron-Cohen, O'riordan, Stone, Jones, & Plaisted, 1999; Happe, 1994). Studies using these measurement tools each have suggested that children develop important social cognitive insights during the latency years.

Limitations of Social Cognitive Measures

Despite these findings, several critical issues have limited research on social cognition measures. First, many theory of mind measures are "paper-and-pencil tasks" in which social scenarios are read by or to participants. These tasks place reading and verbal demands upon children and may be more sensitive to tangential cognitive operations, including attention, executive processes, and verbal comprehension. Researchers have suggested that much of the variability in theory of mind performance amongst children may be attributed to developments in other cognitive functions, particularly in verbal domains (e.g., Liss et al., 2001). Second, theory of mind tests may not correspond directly to social functioning. Participants are seldom made to formulate on-line inferences as in naturalistic social scenarios. Tasks that provide subjects with static vignettes do not provide the same breadth of cues involved in naturalistic interactions. Indeed, several studies have suggested that performance on theory of mind tasks may not always correspond directly with real-life social functioning (Custrini & Feldman, 1989; Leppanen & Hietanen, 2001). Third, few advanced social cognition tasks have been implemented with large samples of typically developing children. Many reported measures document differences between specific populations (e.g., autism, patients with neurological damage) and typically developing children. However, only by examining developmental differences between different age groups of children can researchers truly elucidate typical developmental pathways and the mechanisms underlying these changes. Fourth, few measures have been used in multiple studies, limiting the generalizability of findings and the ability of researchers to identify factors that predict children's performance (e.g., gender, age, verbal ability).

Several additional limitations emerge from studies of emotion recognition measurement. First, emotion recognition skills may be different for distinct emotions.

Neuroscientific research has suggested that distinct neural networks may be associated with the processing of positive and negative emotions (Adolphs, 2002), and that these networks may develop at different rates (Giedd et al., 2003). Rarely have studies examined the development of emotion recognition for distinct emotions separately. Second, children appear to reach adult levels of recognizing basic facial emotions presented in pictures at an early age. However, studies have indicated that although progress in emotion recognition begins to stabilize around the age of 9-12 years in females, males may continue to develop until early adolescence (Harrigan, 1984). Also, these studies may not provide the entire picture because children only view isolated pictures without the complexities of a naturalistic social event. New, more naturalistic measures are important for mapping gender specific and emotion specific pathways of emotion recognition development.

<u>Present Investigation</u>

The present study sought to investigate three measures of social cognition development that differ in their target skills (i.e., advanced theory of mind processing, emotion recognition, intentionality detection) and the cues that they provide (i.e., contextual, facial and other nonverbal gestures, goal-directed movements). Two of these measures (Faux Pas Task and Animations Task) were chosen because they have been used in recently published research, studied multiple times, shown to be sensitive to neuropsychiatric impairments, and reported to be reliable and valid. The Animations task also was chosen because it provides subjects with dynamic stimuli for which they need to track ongoing interactions. It was hypothesized that this measure would be the closest approximate of how children would view real- life social interaction. The third task (Movie Stills) was chosen due to its reported sensitivity to emotion recognition deficits in a patient population (ie., patients with amygdala damage), its capacity to separate facial cues from other nonverbal social gestures, and the complexity of its stimuli. Each of these tasks was administered to a sample of latency-aged typically developing children to examine the measures' sensitivity to social cognitive developments, ability to measure the skills which they intend to measure, and capacity to measure distinct functions. Additionally, factors associated with developmental differences between older and younger children were investigated for each task. Specific goals for each task are discussed below.

Faux Pas (FP) Task

Baron-Cohen et al. (1999) developed the Faux Pas Task (FP) as a measure of children's ability to identify social faux pas, or awkward and inappropriate statements that are potentially offensive without the intent of inflicting harm or causing embarrassment. The FP consists of ten short stories in which two or more characters interact, and one of these characters "says something that they should not have said." An example of a FP item is as follows:

> *Kim helped her mom make an apple pie for her uncle when he came to visit. She carried it out of the kitchen. "I made it just for you," said Kim. "Mmm," replied Uncle Tom, "That looks lovely. I love pies, except for apple, of course!"*

In order to identify such a social blunder, children must comprehend the context surrounding the event, while also recognizing the separate mental states of each character both before and after the target statement. Children also are asked one question regarding the intent of the character committing the faux pas for which they must recognize that the speaker did not maliciously commit the identified inappropriate statement. Therefore, children's ability to correctly respond to FP items requires an integration of multiple variables, understanding of each character's knowledge state, perception of the emotional impact of the target statement, and a comprehension of the intent of the speaker.

Originally, the authors administered the FP to sixty children between 7 and 11 years old and indicated that older children correctly identified more faux pas items than younger children (Baron-Cohen et al., 1999). The authors also reported that typically developing children outperformed IQ-matched children with autism. These findings suggested that the FP was sensitive to theory of mind developments appearing during latency years, as well as theory of mind deficits characteristic of children with autism. This measure has since been used in several studies and been shown to be sensitive to theory of mind impairments in patients with bilateral amygdala damage (Stone et al., 2003), frontotemporal dementia or Alzheimer's disease (Gregory et al., 2002), and nonverbal learning disorder (Dorfman, 2001). Still, only one of these studies (i.e., Dorfman, 2001) investigated the FP test in children, and this study compared only 19 children to a group of severely impaired learning disabled children.

Amongst these studies, several limitations emerge. First, none of the authors adequately investigated the verbal intelligence of their subjects. The original test developers (Baron-Cohen et al., 1999) reported that verbal mental age (VMA) was significantly correlated with children's performance on the task ($r=.52$, $p>.05$), though they do not account for this relationship when examining differences between age groups. Thus, it is possible that improved performance on the FP could be moderated by improvements in understanding more subtle aspects of the stories or acquired skills in articulating final responses. The authors counter this potential criticism by arguing that children responded to a general comprehension question for each story, and that children in the study had to pass a minimum of 7 out of 10 comprehension questions to be included. However, these comprehension questions were somewhat basic, and the authors did not specify how many children failed to correctly answer a sufficient number. The comprehension questions do not account for more subtle verbal nuances and thus may not address more complex verbal and comprehension demands.

A second limitation of these studies is that they do not tease apart the components of theory of mind processing that are developing amongst typical children. The original authors (Baron-Cohen et al., 1999) do not postulate why children who were able to accurately identify one faux pas were not able to identify the remaining items. They indicate that no items proved to be significantly more or less difficult than others. But, the authors do not offer an explanation as to why children who were able to score above chance on the test were not able to correctly respond to all of the items, despite clearly having acquired some capacity to detect inappropriate statements. Two different mechanisms may account for this perplexing finding. First, older children may not have a more developed theory of mind *per se*, but they may be more skillful and consistent in applying this skill to novel circumstances. Such discrepancies have been reported in previous research on language and memory abilities and suggest that children may be competent in a cognitive domain before they are able to consistently apply this competence (Ornstein, Naus & Liberty, 1975). Second, there may be an underlying difference between some of the items or groups of items that allows the targeted statements to be more or less identifiable. Stories differ on several dimensions, including length, placement of the target statement, character responses to the target statement, and identifying words within the faux pas that make the statement more overt (e.g.

“horrible” or “weird”). Therefore, it is possible that the authors’ measure may be tapping developmental changes in theory of mind skill with items that may be differentiated by factors unrelated to children’s ability to infer the intentions and mental contents of others. These questions each were examined in the present study.

Movie Stills Task (MS)

The Movie Stills Task (MS), developed by Adolphs and Tranel (2003), measures individuals’ ability to recognize emotions in complex scenes. This measure includes two series of pictures, each depicting complex social scenes. The two conditions are distinguished from one another only by the presence or absence of faces. Subjects are asked to label the emotion that most closely matches the way that the character in the center of the photo is feeling. By utilizing two sets of pictures, the authors were able to tease out the level of improvement gained from being able to use facial information to recognize emotions.

Adolphs and Tranel (2003) previously reported that subjects with neurological damage to the amygdala, a neural structure associated with processing facial emotions (for a review, see Adolphs & Tranel, 2000), performed significantly worse at identifying emotions from scenes with faces than neurotypical subjects and subjects with non-amygdalar neurological damage. In contrast, patients with amygdala damage did not show impairments in their ability to recognize emotions from scenes with faces erased. These findings suggest that the MS task is sensitive to facial emotion recognition deficits.

This measure may be particularly useful for mapping the development of emotion recognition from facial cues within complex, naturalistic social scenes. Most studies examining facial emotion recognition development control for social cues outside of the face by excluding them. This measure offers a novel method of disentangling facial cues from other social gestures and presents subjects with scenes that more closely resemble real-life social events. However, no prior studies have applied this task to children. By administering this task to children, we may better understand the development of children's ability to recognize different emotion cues and different emotion types.

Animations Task

Abell, Happe and Frith (2000) first developed the Animations task, based on the work of Heider and Simmel (1944), in order to measure children's ability to accurately infer mental states and the intentions underlying behavior. This task depicts geometric shapes moving in a goal-directed manner as if they were interacting. Heider and Simmel's original study indicated that most neurotypical adults naturally ascribe agency and intentionality to these shapes. Studies using versions of this task have demonstrated that non-human primates (Uller & Nicholls, 2000), infants (Gergeley et al., 1995 in Klin), children (Abell, Happe, & Frith, 2000; Bowler & Thomen, 2000; Springer, Meier, & Berry, 1996) and adults (for an extensive review, see Scholl & Tremoulet, 2000) spontaneously attribute "mentalizing" behaviors, social intentions, thoughts, beliefs, and emotions, to the shapes. This measure thus serves as a wonderful resource for investigating theory of mind inferences, emotion recognition, and social processing amongst different populations.

This newer version of the moving shapes task was made more child-friendly, with the shapes being drawn in different colors and the scenes depicting more child-like themes (e.g., playing, follow-the-leader). Unlike the other two tasks in this study, the Animations task provides children with a dynamic scene in which characters interact. This task is more naturalistic than most social cognition measures in that the shapes are moving throughout.

Results from this task and similar tasks have indicated that the performance of older children is superior to younger children. For example, Springer, Meier and Berry (1996) indicated that 5-year-old-children's characterization of geometric shapes moving in a goal-directed manner were more similar than those offered by 3- and 4-year-old children to adult descriptions. Additionally, Abell, Happe, and Frith (2000) demonstrated that 8 year-old typically developing children outperformed mental-age matched children with autism. However, both studies failed to account for children's verbal ability. It is not clear whether superior performance was predicted by children's vocabulary level, ability to describe mentalizing terms, or social cognitive skill. Conclusions from the Springer, Meier, and Berry (1996) study also were limited by their failure to discuss whether the children became more descriptive, more accurate, or both in their narratives. Children's improved ability to describe the behaviors depicted by the geometric figures may be a result of their developing verbal abilities and increased vocabulary, improved intentionality detection, or greater skill in recognizing emotions. These components have not yet been teased apart.

The sensitivity of this test to the developing social abilities of typically developing children has not been examined. Abell, Happe, and Frith (2000) investigated a small sample of 8 year-olds, allowing the authors to report that children make social attributions from shapes but limiting their ability to draw conclusions about the development of this inference-making skill. In a similar study, Bowler and Thommen (2000) utilized a geometric figures task but focused only on the performance of a single age-group and did not investigate complex mental state attributions (e.g., bluffing, deception). Developmental studies are necessary for providing a baseline against which to compare atypical development and for supplying information regarding the utility and specificity of tasks of moving geometric shapes.

Summary

Research has shown that children exhibit critical theory of mind and emotion recognition developments throughout the latency years. However, less is known about the mechanisms by which children improve in their theory of mind processing. Limitations in the research on social cognition measures have hindered progress in this field. Many advanced social cognition measures have been implemented with psychiatric or neurological patients but have not been administered to large samples of typically developing children. Also, when they have been used, researchers have seldom examined individual items of these measures in order to tease out specific variables integral to children's social cognitive development. Finally, children's performance has not always been examined in relation to changes in other cognitive domains. The present study sought to address each of these limitations with three social cognition tasks.

Hypotheses

Hypothesis 1. Based on previous research indicating that the development of theory of mind processing is highly related to emotion recognition development in typical children (Buitelaar & van der Wees, 2003), we predicted that children's performance on each of the three studied social cognition measures would be correlated with one another.

Hypothesis 2. This hypothesis focused on children's performance on the FP task, and contains two related predictions. First, previous studies have indicated that older children performed significantly better than younger children on the FP task (Baron-Cohen, et al., 1999). We thus predicted that chronological age would serve as a significant predictor of performance. Second, studies have indicated that children's performance on theory of mind tasks in which children must comprehend verbal material is modulated by their verbal ability (Happe, 1994). We thus predicted that children's performance would also be significantly predicted by their verbal mental age.

Hypothesis 3. Based on research investigating emotion recognition development, we hypothesized that chronological age would be a significant predictor of children's performance on the MS task. Also, we hypothesized that children would show superior performance on scenes with faces, than scenes without faces. We also hypothesized that, because this task did not place large demands on children's verbal comprehension, that chronological age would be a significant predictor after covarying the effects of verbal mental age.

Hypothesis 4. We hypothesized that older children would describe social scenes by using more mental state references, and also would be more accurate in their descriptions than younger children. We believed that these differences would be evident for complex social scenes in which the shapes moved as if to affect each others' mental states, but not for scenes with random movements. Also, because this task was not dependent on children's ability to understand verbal information, we thought that chronological age would be a significant predictor above and beyond the predictive value of verbal mental age.

Hypotheis 5. Hypothesis 5 was largely exploratory. We sought to investigate whether any of the FP items were more or less difficult than other items. We also wanted to investigate whether length of vignettes, in which portion of the vignette the faux pas occurred, how many characters were present, or whether the faux pas contained a negative connotation or overtly inappropriate statement was related to the difficulty of items.

Hypothesis 6. This hypothesis was largely exploratory as we sought to investigate differential patterns of performance between emotion types on scenes with and scenes without faces.

Method

Participants

Sixty-four children between 6-12 years old (37 male, mean age=112 months, age range=73-148 months) were recruited through three local recreational camps. Three social cognition tasks and one receptive vocabulary test were administered to each child. With the exception of the receptive vocabulary test, the order of tasks was counterbalanced across participants. The vocabulary task always was administered first in case children demonstrated significant verbal impairments that either would confound results or prevent the child from fully comprehending the tasks. This did not occur during the course of the study.

Measures and Procedure

Verbal ability

Verbal ability was measured with the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981). During this test, children must choose from a series of four pictures the option that most closely matches the word given by the examiner. For example, the examiner would read "Point to cat." This test has been normalized on age-referenced groups among a large sample of healthy children and adults between ages 2.5 and 90 years old. Based upon these samples, children's raw scores, or the total number of correct items, may be converted into standardized scores with a mean of 100 and standard deviation of 15, percentile ranks, stanines, and age equivalents. Because verbal mental age (VMA) offered an indication of children's current verbal functioning independent from their chronological age, these values were used for all analyses.

This test was selected for the present study for several reasons. First, unlike many expressive verbal tasks, the PPVT-R does not measure children's knowledge of factual information or social rules. The PPVT-R is a more direct measure of children's knowledge of words. Second, the PPVT-R is less susceptible to underestimates resulting from children becoming anxious or bored during testing. Unlike tasks that require children to verbalize responses, the PPVT-R's only demand is that children point to the appropriate item. The PPVT-R does not require children to explain answers and can be easily administered to younger and more hesitant children (Dunn & Dunn, 1997). Third, the current project included three measures of social cognitive processing, each potentially lasting up to 30 minutes. The PPVT-R is more quickly administered than other verbal tasks, decreasing the likelihood that children would become fatigued. Fourth, this test has been shown to be highly predictive of verbal and overall IQ (Dunn & Dunn, 1997). Finally, previous reports have shown that children's performance on picture vocabulary tasks is highly correlated with their performance on social cognition tasks (e.g., Baron-Cohen, O'riordan, Stone, Jones, & Plaisted, 1999; Happe, 1994). Therefore, by incorporating the PPVT-R into final covariate analyses, we were more likely to be measuring that component of verbal ability that may confound interpretations of social cognition testing.

Psychometric Properties. Studies have indicated that the PPVT-R demonstrates good psychometric properties (Dunn & Dunn, 1981). The test's authors computed alpha coefficients to measure internal consistency. Coefficients ranged from .93-.98 across age groups (median=.95). Split-half reliability coefficients derived from comparing odd- and even-numbered items also were high (range = .86-.97; median = .94). Test-retest reliability analyses indicated that children's performance during an initial administration and then one month later was consistent (median alpha=.92).

The PPVT R also appears to be a valid measure and is highly correlated with other vocabulary test and subtests, as well as IQ tests (for a review, see Williams & Wang, 1997). Correlations with vocabulary and IQ tests for the age groups in the present study ranged from .59 to .76. Also, items were selected from a pool of words on the basis of normative data of a large sample of healthy individuals. Items only were included if their growth curves were congruous with other items on the test, and percentage passing rates for each group indicated that a slightly higher percentage of children for successive age group correctly identified the item (Dunn & Dunn, 1981).

Social Cognition Measures

Faux Pas Task (FP). The Faux Pas task (FP) is a fifteen-item measure in which subjects read vignettes of 2-3 sentences, 10 of which contain a social faux pas (Baron-Cohen, O'riordan, Stone, Jones, & Plaisted, 1999). The remaining five items were matched on length and complexity and were comparable to the faux pas items in every way except that they did not contain a faux pas. These items served as control stories to increase the likelihood that children would not be able to answer questions correctly by assuming that a faux pas was committed. Subjects then were asked four separate questions regarding 1) whether or not a faux pas had been committed, 2) if yes, what the faux pas was, 3) a comprehension question specific to each item, and 4) a *false-belief* question regarding whether or not the speaker knew that their statement would be offensive or embarrassing. The first two questions probed whether or not children detected a faux pas. The third question assessed children's ability to attend to and understand the story. The final question was asked to assess whether children understood that the faux pas was a result of an absence of knowledge or memory, rather than malicious intent. Order of items was counterbalanced across subjects to eliminate novelty or fatigue bias, and control items were evenly dispersed amongst stories containing a faux pas.

Items in which children neither identified the faux pas nor correctly answered the comprehension question were not included in final analyses. Items were scored correct only if each question was answered correctly. However, additional analyses were performed in which a total score (0-2) for faux pas questions was given for each item (not counting a point for comprehension questions and grouping the first two questions into one item). This method provided information regarding whether children were able to correctly identify the faux pas as well as perceive the intention underlying the statement.

Movie Stills (MS). The MS task consists of sixteen black and white pictures from commercial films with faces erased as well as a series of the same pictures with faces included (Adolphs & Tranel, 2003). This task has been reported to be a measure of emotion recognition from facial and nonfacial cues that is sensitive to specific neurological damage (IBID). Stimuli were selected by the test authors based upon the reliability of responses from healthy adults in a pilot study. All stimuli show complex scenes with at least one person displaying a range of simple and complex emotions. The stimuli varied in terms of the particular cues that provided information about the emotion, but each contained multiple cues (e.g., body posture, head posture, direction of eye gaze, hand and arm gestures), including facial expressions.

Subjects are shown each of the 16 photographs with faces erased and asked to match the photo to a list of seven words: happy, sad, angry, afraid, disgusted, surprised, or neutral in terms of how they thought the people in the picture felt. Each word was read and explained to children in child-friendly language matching as closely as possible the definitions found in the Oxford Dictionary, and children then were asked if they understood the meaning of all the words. In the case that they did not, words were re-explained until the child indicated that he or she understood. Following the administration of the first 16 photographs, children were then shown each picture with faces included and given the same instructions.

Children had no time limit on this task. In the case that there were multiple people in the photograph, children were shown which character to focus on. In each case, this was the person who was most directly facing the viewer and judged to occupy the primary role in the scene.

Performance by children was scored in relation to previously acquired healthy adult responses in order to form a continuous score of proficiency. Thus, if 100% of normal subjects called a scene “happy”, then a subject would get a score of 1.0 for choosing the label “happy” and 0.0 for all other choices. However, in the event of more ambiguous stimuli, children’s scores ranged between 0.0 and 1.0 based upon the proportion of healthy adults who labeled each of the possible emotions for that particular photograph. For example, if 50% of subjects labeled a photograph as “surprised,” 40% called it “afraid,” and 10% called it “sad,” a subject would receive a 1.0 for choosing “surprised,” .8 for choosing “afraid,” and .2 for choosing “sad.” Thus, correctness was made a parametric function of the distribution of responses given by healthy adults. In this manner, higher scores represent superior recognition ability. In order to identify the degree to which performance improved when faces were presented relative to the faces erased condition, we subtracted scores on the initial condition (i.e. faces erased) from scores on the latter condition (i.e. faces included).

Animations. Animations were designed previously using a MacroMedia Director 4.0 program for Power Macintosh (Castelli, Frith, Happe & Frith, 2000), and were converted to .avi files and presented with QuickTime Player software. Each vignette consisted of one red triangle and one blue triangle moving around the screen, usually with a blue square enclosure. Three conditions of Animations were shown to children: theory of mind (ToM), goal-directed (GD), and random (RD). The three conditions were matched for length and complexity, with all animations lasting between 34 and 45 seconds. Also, each condition showed enclosures and shape changes (e.g., a triangle squeezing through a gap) at an equal rate in order to ensure that the differences in interpretation were not due to any configural alterations between conditions. Ten vignettes were shown to each subject, including four ToM, three GD and three RD animations. One GD and one RD animation each were shown to children prior to the experiment to familiarize them with the task. Including these practice items resulted in one fewer test item for two conditions (i.e., RD and GD) compared to the ToM condition.

In the RD animations, the triangles did not interact with each other and moved without purpose (e.g., floating in space, bouncing off walls). In the GD sequences, one triangle responded to the other triangle and the two demonstrated a simple interaction in which their movements were contingent upon one another (e.g., one triangle leading the other). For each of these vignettes, there was reciprocal interaction without any indication that either of the triangles was affecting the mental state of the other triangle. In the ToM vignettes, the two triangles interacted in a manner that suggested that one of the triangles was deliberately trying to read or affect the mental state of the other triangle (e.g., one triangle surprising the other, one triangle mocking the other, one triangle persuading the other, one triangle tricking the other). In all three conditions, the triangles moved as if self-propelled. In previous studies, children were given cues as to what they were going to see contingent upon the condition (Abell, Happe and Frith, 2000). In contrast, in the present study, in order to reduce children's bias for attributing social meaning to stimuli, children were not primed for any condition and were merely instructed to "Watch the cartoon."

Following each scene, descriptions were elicited by asking subjects, "What happened in that cartoon?" Feedback was given if necessary (e.g., try to pay attention to the entire cartoon) during the practice items, but not during test items. General encouragement was offered during the experiment proper, and all responses were recorded. Animations were presented in a pseudo-random order and counterbalanced across subjects.

Subjects' responses were scored along three dimensions: intentionality, appropriateness, and length (see Appendix A for scoring guidelines). Intentionality referred to the degree of mental state attribution reported in a given narrative and was scored between 1-5 with 1 representing no intentionality, and 5 being given to those narratives in which descriptions referred to action with the deliberate intent of affecting the other character's mental state. Appropriateness was scored between 0-2 for each narrative with higher scores indicating that children had accurately identified all of the intended actions in the scene, and lower scores indicating that narratives reflected only a minor and non-central aspect of the story or implausible action. Therefore, appropriateness scores were independent of intentionality scores. For example, children received an intentionality score of "0" for the description “The two triangles were bouncing around.” This description also would be scored as a “2” for appropriateness if it referred to a RD vignette.

Finally, narratives were scored for length. In contrast to original scoring guidelines (Abell, Happe & Frith, 2000) in which all narratives with more than 3 clauses were coded as a 3, the number of clauses was scored in a continuous manner. This was done in order to decrease the likelihood that differences would be obscured by a restricted range. Clauses were defined as a subject and a predicate and did not include clauses that were irrelevant to the action of the narrative (e.g. “It looked like” or “and then it ended.”)

Three independent raters performed all scoring for the Animations narratives. In order to calculate both inter-rater and intra-rater reliability, each rater independently scored 5 cases, each containing ten animations, on two separate occasions. The Rater Agreement Index (RAI; Burry-Stock, Shaw, Laurie, & Chissom, 1996) was used to calculate both inter- and intra-rater reliability coefficients for Intentionality and Appropriateness. The RAI measures the degree to which coders agree on their ratings in reference to the possible range of ratings. The index ranges from 0 to 1, with 1 indicating perfect agreement. The basic formula for calculating these values is: $RAI = 1 - (|R_1-R_2| / (I-1))$ where R_1=coder A’s rating, R_2=coder B’s rating, I=the range of the scale. The RAI was selected because, unlike other reliability coefficients, it assesses separate raters’ ability to both score similarly to one another and remain consistent over time. Reliability on Length measures was analyzed using Pearson correlations because there was no finite range of values for this variable.

RAIs for interrater reliability on Intentionality and Appropriateness of responses ranged from .88 to .91 and .89 to .90 respectively. Pearson correlations for length ranged from .87 to .92. Also, RAI’s for intrarater reliability on Intentionality and Appropriateness ranged from .94 to .96 and .91 to .96 respectively, whereas Pearson correlations for length ranged from .89 to .94. Raters demonstrated adequate reliability using the Animations scoring procedures; consequently, inferential statistics were performed.

To test the hypotheses, an average of two raters' scores was used for each response. However, when the scores for the two raters differed by more than one for either Intentionality or Appropriateness, the third rater scored this item and the average of the two scores closest together was used. These scores were used for all analyses. In the case of Length, when two ratings differed by more than 2, then all three raters discussed the item until they could reach a consensus on the number of clauses.

Results

Relationship Between Measures

In order to examine the relationship between the three social cognition measures, we correlated each element of the FP, MS, and Animations tasks (see Table 1). These analyses were performed in order to address whether the measures were sensitive to the same latent construct and to determine whether tasks with static cues (i.e., FP task and MS task) predicted performance on a task with more naturalistic, dynamic interactions.

Two-tailed Pearson Product Moment correlations indicated that scores on the FP tasks were positively correlated with performance on both conditions of the MS measure (with faces condition: $r(60)=.302$, $p=.015$; without faces condition $r(60)=.285$, $p=.027$). Faux Pas scores also were significantly correlated with appropriateness scores for ToM animations $r(60)=.332$, $p=.008$, but not with any other Animations outcome scores. Interestingly, these results suggest that performance on the FP is closely related to children's ability to recognize emotions and to appropriately infer the intentions underlying social interactions. However, FP task scores were not as closely associated with children's increasing tendency to use more mental state language in Animations scenes, or to their accuracy for less complex scenes.

Performance on the MS condition with faces was significantly correlated with intentionality scores on all three Animations conditions (RD: $r(61)=.380$, $p=.002$; GD: $r(61)=.290$, $p=.020$; TOM: $r(61)=.360$, $p=.004$), suggesting that emotion recognition ability is related to the level of intentionality that children attribute to nonverbal, dynamic portrayals of social interactions. But, results suggesting that higher MS scores were predictive of greater intentionality attributions in the RD condition indicate that children may overattribute goals and social intentions. Similar results were found when examining the relationship between the MS condition without faces and intentionality scores for the three Animations conditions (see Table 1). Thus, these measures appear to be related, suggesting that children's developing ability to recognize emotions in complex scenes, whether conveyed through facial expressions or other cues, is highly related to their ability to ascribe intentions to interacting stimuli.

When examining the relationship between emotion recognition abilities and children's ability to accurately describe the most complex of the Animations scenes (i.e., ToM vignettes), we found that scores in both MS conditions were highly correlated with children's Appropriateness scores (with faces: r(61)=.359, p=.004; without faces: r(61)=.425, p=.001; see Table 1 for additional analyses). Relationships amongst MS performance and the GD conditions were more modest and only significant for the without faces condition r(61)=.278, p=.025. In contrast, Appropriateness scores for the RDwere negatively correlated with MS scores in the with faces condition r(61)= -.256, p=.028), and not were not significantly related to the without faces condition r(61)= .027, p= .265. These findings suggested that children's increasing aptitude in recognizing emotions may make them more likely to attribute higher levels of intentionality to Animations scenes, but these attributions are not always accurate, and findings on the RD condition suggest that children may actually be overattributing intentions as they increase in age. Moreover, different patterns of results emerged between the with faces and without faces conditions, indicating that these two conditions may have different levels of sensitivity to emerging emotion recognition skills.

Examination of all three conditions is therefore important for understanding children's level of intentionality detection and the accuracy with which it is applied.

Together, these results indicate that developments in faux pas detection parallel developments in emotion recognition from both faces and other gestures. The FP task also seems to measure developments that parallel advances detected by the Animations task, namely those related to accurately understanding the intentions of objects in the ToM scenes. The relationship between FP scores and Animations scores was not significant for any other condition or score type, indicating that these tasks may be tapping unique constructs, and that only the most complex Animations scenes may be measuring skills that overlap with children's ability to detect faux pas. However, performance on the Animations task was highly correlated with both conditions of the MS task. Different patterns of association emerged between different conditions for both tasks, suggesting that each of these conditions may be useful for understanding different aspects of children's social cognitive development.

Sensitivity of Individual Measures

Following these analyses, we examined the sensitivity of each individual measure to children's developing social cognitive skills. We regressed children's VMA, gender, and chronological age on each of the dependent measures (i.e., scores from the primary measures). Subject characteristics are listed in Table 2. For each model, VMA was used as an indicator of children's verbal ability. This score was derived from children's performance on the PPVT-R and was chosen due to the restricted range of standardized scores.

Three subjects were identified as outliers based upon their VMA. Because of the evenness of the upper end of the distribution of VMA, small increases in raw scores above 300 resulted in large changes in VMA. As a result, three subjects had VMAs more than 2 standard deviations from the mean. Therefore, these subjects were not included in final analyses.

All predictor variables were centered prior to analyses in order to reduce the effects of multicollinearity. In addition, assumptions of normality and homoscedacticity were met for each independent variable. Dummy codes for gender (females=0, males=1) were used in all analyses. Chronological age and VMA were converted into total months and, therefore, treated as continuous variables. This procedure allowed for more statistical power and eliminated the possible confounds of selecting arbitrary cutoff points for separate age groups. Alpha was set at $p<.05$ for all analyses.

Faux Pas Task

We first examined individual results computed by adding children's scores for identifying the faux pas (i.e., question #2 for each item) and false beliefs questions (i.e., question #4 for each item) across all items. Therefore, children's scores ranged from 0 to 2 for each item, and the total range was 0-20 for the FP task. A hierarchical linear regression was conducted, using VMA, gender, chronological age, and an age by gender interaction term, respectively. Comprehension questions were not included in the total scores but were examined in order to eliminate cases in which children were unable to identify a faux pas because they did not understand or attend to a given story. If children did not correctly answer the comprehension question but did get the remaining questions correct, his or her score was still used in final analyses. This case-wise deletion process eliminated four items from a total of three subjects.

Results of examining VMA, gender, chronological age and the interaction term as predictors of FP score were not significant $F(3, 58)=2.335$, $p=.067$, but VMA was significantly correlated with total items correct r (59)=.258, p=.045. These results suggested that this measure was not sensitive to theory of mind differences above and beyond developments in verbal ability.

Item Analysis

A second purpose in investigating the FP task was to examine why children who correctly identified some faux pas were not able to label each remaining faux pas. To study this, we examined whether any of the 10 faux pas items were answered correctly either more or less often than other items. In fact, we found that there was a large range between items in the proportion of children who responded correctly (.35-.93) (Table 3). These results indicated that items likely differed in some important way(s).

We thus correlated the proportion of children who responded correctly to each item with several variables. These variables included the following: 1) number of words, 2) number of sentences, 3) number of characters in the story, 4) number of verbal exchanges, 5) placement of the faux pas within the story, 6) whether or not one of the characters responded to or redirected attention away from the faux pas, and 7) whether or not the faux pas contained some negative connotation, as if intended to offend someone (e.g., "those curtains are horrible," "would you like to hear my joke about sick people?"). The only significant relationship found was that between the proportion of children responding correctly to an item and whether or not that item contained a negative connotation. Analyses suggested that items with a negative connotation were answered correctly more often than those without a negative connotation (t (10)= -2.785, p=.014). Items with a negative connotation were characterized by their inclusion of a statement that might be construed as inappropriate regardless of what is known about the characters' mental state(s) (i.e., "That new boy looks weird"). Children could have used the negative connotation to surmise that something was said that "should not have been said" without their knowing precisely that the affect of the statement was that someone's feelings were hurt. Moreover, one item in which a character committed a faux pas by expressing empathy towards another character (i.e., "I'm sorry that your story did not win") without knowing that the character had not yet heard the bad news was answered significantly less often than any of the remaining items $\chi^2(1, N = 10)=9.284$, $p<.0025$. This finding suggested that the manner in which the faux pas was committed may have impacted whether or not children perceived it as an inappropriate statement. These results indicated that the connotations of the statements themselves, as opposed to their social context, may modulate whether or not children identify faux pas as inappropriate.

Finally, we analyzed whether older children performed better or equal to younger children according to the level of difficulty of the items. In order to investigate this relationship, we compared the oldest ($\geq$ 11 years) and youngest ($\leq$ 7 years) children on both easy (i.e., those with a negative phrase or word) and difficult (i.e., those remaining) items. Results indicated that older children performed significantly better than younger children on items with a negative connotation $t(22)= 2.256$, $p=.038$ but not on other items $t(22)=1.214$, $p=.238$. These findings indicated that older children were better able than younger children to identify faux pas only when the faux pas contained a negative connotation.

Movie Stills Task

Next we examined the sensitivity of the MS task to developments in emotion recognition skill. Children's scores for both the "with faces" and "without faces" conditions were analyzed as a parametric function of the distribution of scores reported by healthy adults. The test developers (Adolphs & Tranel) previously collected these results and aided in the computation of children's scores. A paired-samples t-test was computed to compare mean performance across emotions for the two conditions. As expected, results indicated that subjects more accurately identified emotions in pictures with faces than those without faces $t(63)=9.374$, $p<.001$.

Based upon these findings, we analyzed the two conditions separately. Hierarchical multiple regressions were conducted. VMA, gender, and chronological age were entered as independent predictors for each condition. MS scores were the dependent variables. Results suggested that the overall model was a highly significant predictor of performance on the MS task when faces were erased $F(3, 61)=5.987$, $p<.001$, and that chronological age $t(61)=2.615$, $p=.011$ accounted for a significant proportion of the variance above and beyond the variance accounted for by VMA and gender. Neither VMA $t(61) = .833$, $p=.409$ nor gender $t(61) = -1.609$, $p=.113$ were significant predictors. However, because previous findings had indicated that females outperform males on emotion recognition tasks during latency years, we performed a Mann-Whitney U test to account for uneven sample sizes of males and females and found that females showed superior performance to males ($U=297$, $p=.030$).

The overall model also was significant for performance on the MS with faces condition $F(3, 61)=5.291$, $p=.001$. Post-hoc analyses indicated that the effect of gender was significant after covarying VMA $t(61)=-2.331$, $p=.023$, suggesting that females performed better than males. Also, the effect of chronological age was significant after covarying VMA and gender $t(61)=2.626$, $p=.011$, suggesting that older children performed better than younger children.

Finally, a difference score was calculated to examine the level to which the presence of facial cues improved children's performance. The overall model was not significant for the difference score between conditions F(3,61)=.255, p=.950.

Item analysis

Because studies previously have shown that children's skill in recognizing emotions differs across emotions, we investigated each emotion separately. The present task contained two "happy," five "afraid," four "angry," three "sad," one "surprised," and one "neutral" scene for each condition. Mean scores for each of these emotions are listed in Table 4.

Results revealed a differential pattern between basic (i.e., happy, sad) and more complex emotions (i.e., surprised, angry, afraid). First, the overall hierarchical model testing VMA, gender, chronological age and a chronological age x gender interaction was a significant predictor when faces were erased for happy scenes F(3,61)=3.071, p=.035 and approached significance for sad scenes F(3,61)=3.140, p=.051. Investigation of both models indicated that chronological age was a significant predictor for happy scenes t(61)=2.798, p=.007, but not for sad scenes t(61)=1.577, p=.120. The model was not a significant predictor for any of the complex emotions or for neutral scenes. These findings suggested that the MS task only is sensitive to developments in identifying happy and sad faces, or that children do not show significant developments in identifying complex emotions or neutral expressions within complex scenes during this age period.

Second, results for scenes with faces revealed that the model was significant only for sad scenes F(3,61)=3.989, p=.012. Further analyses indicated that this significance was driven by the effect of gender t(61)=-2.189, p=.033, with females outperforming males. In contrast, chronological age was a significant predictor of performance on sad scenes only when an interaction term is not considered t(61)=2.440, p=.018. These results suggested that low power (kappa=.105) could have had an impact on our ability to detect significance when an interaction term was added to the equation. Further, low power also could have biased results as there were relatively few pictures in each emotion category. However, the model was not significant for the two emotions that appeared most frequently (i.e., afraid, angry) in either condition.

<u>Animations Task</u>

Last, we examined the sensitivity of the Animations task. This measure was hypothesized to be the most complex for children because it presented children with dynamic interactions in which they were forced to track the mental states of the moving shapes. Additionally, children's performance scores were based both on their ability to infer mental states and on their ability to recognize emotions and detect the intentions underlying the action.

We first compared children's performance across conditions (see Table 5). Analyses revealed that conditions differed in terms of intentionality, appropriateness, and length. Each difference was in the expected direction.

Because children's responses differed significantly for each condition, we analyzed ToM, GD, and RD responses separately. As in previous analyses, VMA, gender, and chronological age comprised the independent predictors along with an age x gender interaction term in a hierarchical equation. Results suggested that the model was significant for intentionality scores for the RD conditions F(3, 59)=3.685, p=.010. Post-hoc analyses indicated that there was a significant interaction of gender and chronological age t(59)=-2.747, p=.008 in which age was not a significant predictor for females, but was significant for males (Figure 1). Older males ascribed less intentionality than younger males to RD vignettes, though there was no difference between older and younger females.

Finally, the model was significant for intentionality in the ToM condition $F(3,60)=3.866$, $p=.008$. Post-hoc analyses revealed that age accounted for a significant proportion of the variance in intentionality after covarying VMA and gender $t(60)=2.251$, $p=.028$. These results indicated that older children consistently ascribed greater intentionality than younger children to ToM vignettes. In contrast, the model was not significant for intentionality in the GD condition $F(3,60)=1.312$, $p=.277$.

We then analyzed the accuracy of children's descriptions. The model was not significant for appropriateness on the RD $F(3,64)=.588$, $p=.673$ or GD conditions $F(3,64)=2.168$, $p=.084$. In contrast, the model was significant for the ToM condition $F(3,64)=7.970$, $p<.001$. Post-hoc analyses revealed that the effect of chronological age was significant above and beyond the influence of VMA and gender $t(64)=2.332$, $p=.023$, and that the effect of gender approached significance $t(64)= -1.749$, $p=.086$. For length, the model was not significant for the RD $F(3,64)=.760$, $p=.555$, GD $F(3,64)=.092$, $p=.985$, or ToM conditions $F(3,64)=.456$, $p=.768$.

Discussion

The primary goal of this study was to examine three advanced social cognition measures and their sensitivity to the important social cognitive developments that occur during latency years. Further, we wished to investigate the specificity of these tasks by examining their relationship with one another. Finally, we sought to examine variables that potentially modulated children's performance on these measures.

Relationship Between Social Cognition Tasks

Consistent with Hypothesis 1, our results suggested that all three social cognitive measures used in the present study are sensitive to subtle developmental changes within the latency years and that aspects of cognitive development tapped by each of the measures are correlated with each other. The correlations between the three social cognitive measures used in the present study suggest that these measures are sensitive to a variety of related social cognitive skills that may reflect a similar latent construct. Earlier research had suggested that theory of mind and emotion recognition functions are integrated in typically developing children (Buitelaar & van der Wees, 1997). Consistent with this, our analyses revealed that typically-developing children's ability to recognize faux pas was highly correlated with their skill in recognizing emotions from complex social scenes. Children's ability to identify faux pas also was related to their skill at appropriately recognizing the intentions and mental states of moving figures, but only when scenes were complex and conveyed the intention of one character attempting to affect the mental state of another character (ToM condition). In contrast, FP task performance was not related to the level of intentionality ascribed to moving shapes or the accuracy with which it was applied when interactions conveyed “less intentional” behavior. These conditions thus may not be sensitive to social cognitive functions necessary for identifying faux pas.

Children’s performance on each of the Animations conditions was more strongly associated with their performance on the MS task, though the pattern of significant correlations differed by score type and condition. For instance, the level of intentionality characterizing children’s responses for each condition of the Animations task were positively correlated with their performance on both conditions of the MS task. However, the appropriateness of children’s responses for the RD condition of the Animations task were negatively correlated with their performance on the MS task. Moreover, this negative correlation only emerged with the with faces condition of the MS task. These results likely reflect children’s tendency to overattribute intentions based on expectations. Children with greater emotion recognition skill may comprehend a greater array of possible intentions, but they may overgeneralize this knowledge in naturalistic interactions. These findings have not yet been explored in adults, but may parallel research indicating that adults do respond to prompts that behavior is intentional, despite an absence of evidence that an action truly is goal-directed (Heider & Simmel, 1944). Also, the two conditions of the MS task may be tapping into separable, though highly correlated skills. The finding that RD appropriateness scores were inversely correlated with scores from the with faces condition but not with scores from the without faces condition may indicate that the former condition is more encompassing, and thus able to detect a greater range of developments, some of which are related to children’s tendency to overattribute intentions.

The Animations task was incorporated into this study because it appeared to provide a more naturalistic measure of children's inference making and emotion recognition abilities. Our analyses indicate that children's emotion recognition skill is related to their performance on the Animations task. In contrast, children's performance on the FP task was not related to their ability to infer intentions on the Animations task. This finding is striking because both of these measures purport to study the development of mental state inferences, or theory of mind processing. The finding that FP task performance is not significantly related to Animations task intentionality detection indicates that the former task may not be generalizable to more naturalistic social interactions. Instead, the FP task may be assessing other cognitive skills, such as verbal ability, reasoning skill, and the ability to identify social rules. However, these cognitive processes do not necessitate a theory of mind. Research investigating the relationship between FP Task performance and other cognitive skills as well as everyday social functioning would be helpful for elucidating the validity of this measure.

Faux Pas Task

Findings from the present investigation did not support the first part of Hypothesis 2 and instead suggested that children's chronological age was not a significant predictor of their performance on the FP task. These results differed from those reported by the original test developers (Baron-Cohen, O'riordan, Stone, Jones & Plaisted, 1999). Several explanations could be offered for this difference. First, our findings supported the second portion of Hypothesis 2 and did suggest that children's performance on the FP task was significantly related to their VMA. Baron-Cohen et al. (1999) also reported that FP task performance was related to VMA in their typically developing sample, but the authors did not account for this relationship in their final analyses. Rather, to address issues regarding the role of verbal ability in completing the FP task, Baron-Cohen and colleagues included children only if they were able to pass 7 out of 10 comprehension questions. The authors did not report how many children this criteria excluded, but in the present investigation, we did not find any children who failed more than one comprehension question. Thus, the comprehension questions may have been too basic to measure children's verbal understanding of more subtle or sophisticated nuances within the items.

Baron-Cohen and colleagues (1999) also suggested that a significant relationship between verbal ability and faux pas performance is expected because both cognitive operations develop throughout childhood. The authors concluded that FP task performance should be correlated with "any other ability that increases with age" (p.411). In contrast, the authors reported that FP task performance was not related to nonverbal IQ, and our results suggested that faux pas performance was not significantly correlated with children's ability to detect intentions of goal-directed objects, which we did find to be a cognitive skill that develops during childhood. The previous authors may not have adequately accounted for the verbal demands of identifying faux pas within the story, while our analyses addressed this limitation and suggested that verbal ability may account for a significant proportion of the developmental differences in faux pas performance.

An additional reason that our results may have differed from Baron-Cohen et al. (1999) is that we characterized chronological age as a continuous variable, whereas the test developers compared three discrete age groups (i.e., 7- vs. 9- vs. 11-year-olds) and defined age as a categorical variable. We chose to characterize age as a continuous variable to eliminate any arbitrary cutoff points, as children develop at different rates. However, it is possible that by only including participants in disparate age groups, Baron-Cohen and colleagues were better able to detect significant differences between participants further apart in age. This seems unlikely because our analyses revealed that children in the upper and lower end of our age range did not differ significantly in FP task performance when VMA was considered as a covariate.

Movie Stills Task

The present study is the first known research to investigate the sensitivity of the Movie Stills (MS) Task to emotion recognition developments in children. This task is an advancement over previous emotion recognition measures for two reasons: 1) children are shown complex social scenes that are more similar to naturalistic social events, and 2) children view two conditions differentiated only by the absence or presence of faces, allowing researchers to cleanly isolate the role of facial displays in children's emotion detection skill.

Findings from the present investigation confirmed Hypothesis 3 and suggested that the MS Task is sensitive to children's developing ability to recognize emotions from both facial and non-facial cues. Children's performance when faces were included improved with age, as did their performance on scenes without faces. A difference score computed by subtracting performance on the without faces from performance on the with faces condition was not significantly related to chronological age. These results indicated that this measure did not detect any age-related changes in facial emotion recognition skill in typically developing latency aged children.

Older children did perform better than younger children on scenes without faces, indicating that children become better able to integrate a variety of non-facial social cues into their percepts. These findings remained significant after partialing out the effects of verbal ability. The above results are consistent with studies indicating that the recognition of nonverbal emotional expressions undergoes a major development during the latency years (McCluskey & Albas, 1981; Philippot & Feldman, 1990). Findings from the present investigation suggested that females performed better than males on both scenes with and scenes without faces. These results are consistent with previous reports documenting improved facial affect recognition in females (for a review, see McClure, 2000), but our finding that females outperform males when faces were erased suggests that gender differences in emotion recognition skill at this age may be due to females' superiority in interpreting a multitude of facial and non-facial social cues .

Animations Task

Analyses of the Animations task supported our predictions in Hypothesis 4 and revealed that this measure was sensitive to developments in the ability to infer the intentions of goal-directed objects. Unlike previous investigations applying this task to children (i.e., Abell, Happe, & Frith, 2000; Montgomery & Montgomery, 1999), our study highlighted this task's ability to measure the development of intentionality detection in a large group of children who already had passed the age at which most children understand that others have separate mental states. We also found that age was a significant predictor of performance after accounting for the effects of verbal developments. This measure thus offers a useful tool for examining social cognition and theory of mind developments with more naturalistic stimuli. Children viewed dynamic interactions which, while being impoverished, still provided multiple dynamic social cues to convey intentions, emotions, and mental states. This task thus may be more representative of naturalistic social functioning and development than static laboratory measures which do not provide children with multiple, visual social cues, or demand that they track ongoing behavioral exchanges.

A goal of the present study was to investigate whether the three Animations conditions were sensitive to developmental changes in the degree to which participants ascribed intentionality to the moving shapes and the accuracy with which they described the interactions between the shapes. First, findings indicated that children's age predicted the level of intentionality observed within the ToM vignettes but not within the other two conditions. These results suggested that older children's descriptions of observed interactions did not contain more social descriptors than younger children's for the two conditions that either were not intended to portray social interaction (RD) or depicted more simple interaction (GD). However, in the ToM condition, older children appeared better able to understand other's mental states and subtle changes in their mental states than were younger children. An example of this difference in complexity was seen in two similar vignettes. Age did not predict level of intentionality within children's descriptions for a scenario in which two triangles played follow-the-leader. In contrast, when the triangles engaged in largely the same movements, but the scenario also depicted some "mocking" behavior performed by the triangle that was following, older children were better able to understand the intent of the mocking triangle and the effects of this behavior on the leading triangle. Therefore, only the ToM condition appeared to be challenging enough to reveal developmental changes.

One interesting finding did emerge regarding the RD animations. Analyses revealed a significant gender x age interaction for intentionalitysuggesting that younger males ascribed greater intentionality to the RD condition than older males. There was no significant effect for females. These findings indicated that younger males in our sample may not yet be able to distinguish intentional from unintentional behaviors. Findings reported by Carpendale and Chandler (1996) suggested that males develop an understanding of the intentions underlying social behaviors at a later age than do females. Our results supported Carpendale and Chandler's work. Younger, latency-aged males may not apply their knowledge of intentionality with the same level of specificity that older latency-aged males do.

Findings also indicated that children's age predicted the accuracy with which they described the vignettes. Older children in our sample were better able to accurately narrate the action in ToM and GD vignettes than younger children. Moreover, the superior performance of older children existed above and beyond their verbal skill. Age effects likely were not present in the RD condition given the relative simplicity of these vignettes and ceiling effects. In the ToM and GD conditions, older children were better able to integrate action from throughout the scenario in order to describe the central purpose of the vignette. Younger children tended to either miss the central point or become overly focused on a minor aspect of the story. Older children appeared to have developed a better "central coherence" or ability to organize critical aspects into a meaningful whole. Measures that do not provide such complexity of dynamic action may not be sensitive to this developmental change. Moreover, this superior performance existed above and beyond children's verbal skill.

<u>Individual itema nalyses</u>

<u>Faux Pas Task</u>

To address Hypothesis 5, we examined individual items of the FP Task. While results suggested that overall, the FP task is not sensitive to age differences in identifying faux pas, our investigation of individual task items did reveal an important finding. If the item contained a negatively valenced statement (e.g, "he looks weird") or word (e.g., "horrible"), both older and younger children were significantly better at identifying faux pas than they were on items that did not include a negative statement or word. Additionally, older children performed significantly better than younger children at identifying faux pas in items with a negative valence, but not in items without.

These results suggested that the FP task is sensitive to children's developing capacity to detect socially inappropriate, or potentially offensive statements, rather than faux pas, *per se*. The overt nature of the inappropriate statement appears to moderate children's skill in deciphering whether or not an individual has committed a social blunder. These negative words or phrases may cue children that something has been said that should not have been said. However, such cues appear to decrease the demands on children's theory of mind as they mark the inappropriate statement without children having to track the mental contents of each character.

Another explanation for why children may have performed better on items with a negative valence is related to the nature of the task instructions. The present study adhered to the guidelines reported by the test developers who suggested that the most child-friendly way of querying subjects' understanding of faux pas is to ask, "Did any of the characters say something that they should not have said?" (Baron-Cohen et al., 1999). This phrasing could pull for faux pas as well as statements that may be intentionally offensive. In other words, children may decide that an individual "should not have said" something for a variety of reasons, including because it was unintentionally hurtful, because it was malicious, or because it violates a social rule. In contrast to other social blunders, faux pas are defined by the test authors as “when a speaker says something without considering if it is something that the listener might not want to hear or know, and which typically has negative consequences that the speaker never intended” (Baron-Cohen, O’riordan, Stone, Jones, & Plaisted, 1999, p. 408). Inherent in identifying faux pas then is having a theory of mind, or an understanding of what the speaker intends to convey and what the listener “might not want to hear or know.” Findings from the present investigation suggested that as children age, they may not improve in their ability to track mental states; rather they may become more adept at attending to negative statements. Because age also was a predictor of children’s skill in identifying the intention of speakers committing a faux pas, it is reasonable to hypothesize that the manner in which faux pas were probed may have driven our findings that items with a negative valence were more identifiable for children.

Movie Stills Task

For Hypothesis 6, we examined recognition of each individual emotion within the MS task. Results from the present investigation indicating that latency-aged children show more improvement in their ability to identify happy and sad emotions in comparison to other emotions is surprising given the research findings that recognition of happy and sad emotions develops rapidly during early childhood. Researchers have suggested that happy and sad are the first emotions that infants are able to recognize. The present findings suggest that despite children’s ability to recognize these emotions early in ontogeny, children in middle childhood still appear to improve in their overall recognition of happy and sad displays embedded within complex social scenes.

We found it surprising that age was not a significant predictor of children's ability to identify the other emotions used in the MS task in either the with or without faces conditions.. Perhaps, affective displays indicating anger, fear, or surprise are more contingent upon an event or interaction and thus require a greater understanding of context than a static picture can convey. It also could be the case that children could not identify more complex emotions because the scenes were too ambiguous. Indeed, the emotions in many scenes were not agreed upon by all the healthy adults in the test developers' original sample. Thus, subtle differences between separate but similar emotions (e.g., fear and anger) may not be detectable in naturalistic scenes without the benefit of some contextual information. Future research may address this issue by testing the minimal amounts of information necessary to convey specific emotions with greater clarity.

Summary

Together these findings highlight the importance of studying individual social cognition measures and their sensitivity to a range of cognitive developments. Two of the tasks studied here (i.e., Movie Stills Task and Animations) were shown to be sensitive to social cognitive developments in latency-aged children, but the FP task was found to be more highly associated with children's developing verbal ability. The finding that chronological age was predictive of performance on the MS and Animations tasks above and beyond verbal ability indicates that they are sensitive to the social cognitive function that they intend to measure. In contrast, findings regarding the FP task suggest that developmental differences may be more related to children's developing verbal skill than their developing theory of mind. Moreover, the individual items within this task appeared to differ in terms of their difficulty, and their level of difficulty was moderated by whether or not they contained a cue that could be identified as inappropriate regardless of whether it was a faux pas. Future research applying the FP task should consider each of these variables when interpreting results.

More broadly, the findings reported here suggest that measuring social cognitive development during the latency years must always consider developments in other cognitive functions, differential patterns of development between genders, the complexity of the social information that is presented, and potential differences between individual items. Additionally, important social cognitive developments emerge in this age period that may be associated with one another or may be somewhat independent. It will be useful to disentangle the associations between separate cognitive functions so that interpretations may become more precise and meaningful. It also will be important to continue to examine the validity and sensitivity of social cognition measures prior to their being applied in comparison studies with atypical populations. Such reports may be further obscuring the social cognitive differences seen in these abnormal populations by misinterpreting the essence of what these tasks are measuring.

Limitations and Future Directions

Several limitations of this study should be addressed in future research. First, this study investigated laboratory tasks of social cognitive functioning that may or may not correspond directly with social functioning or theory of mind processing in naturalistic situations. We attempted to use three tasks that are more naturalistic than traditional measures. However, each of these tasks eliminates several variables and demands present in real-life social situations. For instance, our laboratory tasks offered explicit instructions for children to attend to the stimuli. This direction is not offered in real-life social interactions, and thus children may attend to different aspects of the interaction. Also, tasks were not timed and may have allowed children to develop alternative strategies for solving these problems (Bowler, 1992). Real-life social interactions provide changing information and response demands within an appropriate time-frame. Thus, future investigations may shed light on the relationship between performance on laboratory tasks assessing social cognitive development and more naturalistic applications.

Second, this study included a relatively small, homogenous sample of children. Most children in our study were Caucasian, the children of university professors, and from middle to upper-middle class families. As a result, our sample may not be representative of the larger population.

Additionally, our sample size may limit our interpretations. Several findings approached significance, and relationships examined for separate age groups were hindered by low power. Also, the greater number of males than females may have made interactions difficult to detect. However, only one significant interaction was detected, and no other interactions approached significance. Moreover, several main effects of gender were uncovered in final analyses.

Finally, our use of a receptive vocabulary test to approximate verbal ability may have limited our capacity to account for a greater breadth of verbal functions. While our results suggested that we were able to capture important and related verbal developments, a broader battery of verbal tests may be useful for disentangling these skills from those necessary for sociocognitive performance. Also, other related cognitive processes (e.g., nonverbal IQ, executive functions) will be important to consider in future investigations. Research has more strongly implicated the role of verbal ability in performance on social cognitive measures, but theoretical and some empirical reports also have highlighted a role for these other cognitive domains in social cognitive performance. Although such investigations may warrant longer test batteries, they may begin to tease apart developments and relationships within these different domains.

Conclusion

Overall, our study highlights some of the strengths and weaknesses of three advanced social cognitive measures. Results provided evidence for the sensitivity of two of these measures to the developments in social cognitive functioning emerging during latency years. In contrast, the FP task does not appear to be as sensitive to these social developments, and children's performance appears to be modulated by their verbal ability and the difficulty of items as defined by whether or not they conveyed a negative connotation. Additional variables, including gender, complexity, and the type of emotion/interaction presented further modulated performance on all measures. These variables should be considered in the development of future social cognitive measures as well as in interpretations drawn from current tasks. Our study also highlighted that skills across different domains of social cognitive processing appear to be highly related to each other in typically developing children. Examination of the relationship between these abilities may shed light on underlying impairments mediating social skill deficits in disorders, such as autism, that have been hypothesized to be "disconnect" syndromes in which children are not able to integrate multiple social cues into their percepts.

Appendix A

Animations Task Scoring Guidelines

Scoring verbal descriptions: The verbal descriptions given after each animation were transcribed verbatim and coded in terms of three different dimensions:

Intentionality Score (0 to 5). The Intentionality score reflects the use of mental state terms. The degree of attribution of mental states to the triangles (agents) of the animations is calculated by analysing the content of each description given by the subjects. In the effort to control as much as possible the use of subjective methods in interpreting someone else's language, terminology, idioms and so forth, the analysis is conducted exclusively on the type of verb contained in each sentence used to describe the triangles' actions.

The degree of intentionality reflects in each action is measured with a numerical scale from zero to five. The scale was created explicitly for this task and the type of stimuli that subjects had to describe somehow influenced it. Thus, it is clear that a measure applied to verbal descriptions concerning two agents interacting along a complex motion pattern, cannot and should not reflect the complexity of mental states appreciation contained in any language describing any social context.

In developing the score, the "intentionality ladder" came into shape, with an agent moving upwards, appreciating step by step both actions, and mental states of another agent. At the bottom of the ladder, where there is no appreciation of another agent, nor actions or mental states **(score=0)**, the agent acts with no intention, and no interaction, randomly, e.g. "moving around", or "floating". A further step up in the ladder **(score=1)**, the agent acts with a purpose, a goal, with no interaction with another agent, e.g. "walking", or "swimming". The following step up **(score=2)**, is when the agent acts with a purpose with another agent, e.g. "fighting" or "following": the actions of the two agents are parallel in time. A further step up (**score=3)** is when the agent not only interacts with another agent but acts in response to the other's action, e.g. "chasing", or "restraining": the actions of the two agents are sequential in time. Finally, the two steps at the top of the ladder concern the agent's appreciation of mental states. The lower step **(score=4)** is when the agent acts in response to a mental state, e.g. "arguing", "wanting" or "encouraging". The upper step **(score=5)** is when the agent acts with the goal of affecting or manipulating the other agent's mental states, e.g. "pretending", "deceiving" or "coaxing".

Appropriateness score (0-3)

The Appropriateness score measured the understanding of the event depicted in the animations, as intended by the designers. The score, ranging from zero to three, was based on the underlying script for each animation. Details of criteria for rating the appropriateness of each animation are given in Appendix 4D. The degree of appropriate description of the animation was calculated by analysing the agents' actions and interactions. For example, an appropriate description (score=3) for the animation where the big triangle persuades the little one to go out, need to convey the idea of little triangle's reluctance to go out *and* big triangle's attempts to get the little one out, e.g. "persuading" or "coaxing". A less appropriate description (score=2) would focus on one aspect of the story or one character only. e.g. little one doesn't want to go out; or, big one is pushing little one to go out. An inappropriate description (score=1) concerns actions that do not relate to the events or relate to a very minor aspect of the sequence only, e.g. "the two triangles didn't like each other". Finally, when the subject did not provide any description, the score was zero.

Length Score

Scores reflected the number of clauses within subjects' descriptions. Clauses were defined as a subject and a predicate.

References

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

Adolphs, R. (2002). Neural mechanisms for recognition of emotion. *Current Opinion in Neurobiology, 12,* 169-177.

Adolphs, R., & Tranel, D. (2003). Amygdala damage impairs emotion recognition from scenes only when they contain facial expressions. *Neuropsychologia, 1603,* 1-9.

Adolphs R., and Tranel, D. (2000): Emotion Recognition and the Human Amygdala. In, JP Aggleton (Ed.), *The Amygdala: A Functional Analysis*. New York: Oxford University Press.

Andrews, G., Halford, G.S., Bunch, K.M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74,* 1476-1499.

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

Baron-Cohen, S., O'riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger Syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 29,* 407-419.

Baron-Cohen, S., Ring, H.A., Bullmore, E.T., Wheelwright, S., Ashwin, C., & Williams, S.C.R. (2000). The amygdala theory of autism. *Neuroscience and Behavioral Reviews, 24,* 355-364.

Block, J.H. (1983). Differential premises arising from differential socialization of the sexes: some conjectures. *Child Development, 54,* 1335-1354.

Bornstein, M. & Arterberry, M. (2003). Recognition, discrimination and categorization of smiling by 5-month-old infants. *Developmental Science, 6,* 585-599.

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

Bowler, D., & Thomen, E. (2000). Attribution of mechanical and social causality to animated displays by children with autism. *Autism, 4,* 147-171.

Buitelaar, J.K. & van der Wees, M. (1997). Are deficits in the decoding of affective cues and in mentalizing abilities independent? *Journal of Autism and Developmental Disorders, 27,* 539-556.

Burry-Stock, J.A., Shaw, D.G., Laurie, C., & Chissom, B.S. (1996). Rater agreement indexes for performance assessment. *Educational and Psychological Measurement, 56,* 251-262.

Carpendale, J.I. & Chandler, M.J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67,* 1686-1706.

Castelli, F., Frith, C., Happe, H., & Frith, U. (2000). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125,* 1839-1849.

Crick, N.R., & Dodge, K.A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74-101.

Custrini, R.J., & Feldman, R.S. (1989). Children's social competence and nonverbal encoding and decoding of emotions. *Journal of Clinical Child Psychology, 18,* 336-342.

De Sonneville, L.M.J., Vershoor, C.A., Njiokiktjie, C., Veld, V.O., Toorenaa, N., & Vranken, M. (2002). Facial identity and facial emotions: Speed, accuracy and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology, 24,* 200-213.

Dodge, K.A., & Pettit, G.S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39,* 349-371.

Dorfman, C.M. (2001). Social language and theory of mind in children with nonverbal learning disability. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62,* 574.

Dunham, P., Dunham, F., Tran, S., & Akhtar, N. (1991). The nonreciprocating robot: Effects on verbal discourse, social play, and social referencing at two years of age. *Child Development, 62,* 1489-1502.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology, 23,* 132-139.

Dunn, L.M., & Dunn, L.M. (1981). Peabody Picture Vocabulary Test-Revised: Manual for Forms L and M. Circle Pines, MN: American Guidance Service.

Dunn, L.M., & Dunn, L.M. (1997). Peabody Picture Vocabulary Test-Third Edition: Examiner's Manual for Form IIIA and IIIB. Circle Pines, MN: American Guidance Service.

Flavell, J.H. (1999). Cognitive Development: children's knowledge about the mind. *Annual Review of Psychology, 50,* 21-45.

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

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

Gergeley, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56,* 165-193.

Giedd, J.N., Vaituzis, C., Hamburger, S.D., Lange, N., Rajapakse, C., Kaysen, D., Vauss, Y.C., & Rapoport, J.L. (2003). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4-18 years. *The Journal of Comparative Neurology, 366,* 223-230.

Gregory, C., Lough, S., Stone, V.E., Erzinclioglu, S., Martin, L., Baron-Cohen, S. (2002). Theory of mind in frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain, 125,* 752-764.

Halford, G.S., Wilson, W.H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, cognitive psychology. *Behavioral and Brain Sciences, 21,* 803-831.

Happe, F. (1994). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66,* 843-855.

Harrigan, J.A. (1984). The effect of task order on children's identification of facial expressions. *Motivation and Emotion, 8,* 157-169.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57,* 243-259.

Iannotti, R.J. (1985). Naturalistic and structured assessments of prosocial behavior in preschool children: The influence of empathy and perspective taking. *Developmental Psychology, 21,* 46-55.

Joseph, R. (1992). The limbic system. Emotion, laterality, and unconscious mind. In R. Joseph (Ed.), *Neuropsychology, neuropsychiatry, and behavioral neurology (pp. 87-107).* New York: Plenum Press.

Leppanen, J.M., & Hietanen, J.K. (2001). Emotion recognition and social adjustment in school-aged girls and boys. *Scandinavian Journal of Psychology, 42,* 429-435.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42,* 261-270.

McClure, E.B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126,* 424-453.

McCluskey, K.W., & Albas, D.C. (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *International Journal of Psychology, 16,* 119-132.

McGhee, P.E. (1979). Humor. Its origin and development. As cited in Carpendale, J.I. & Chandler, M.J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67,* 1686-1706.

Morton, J. & Johnson, M.H. CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review, 98,* 164-181.

Odom, R.D., & Lemond, C.M. (1972). Developmental differences in the perception and production of facial expressions. *Child Development, 43,* 359-369.

Ornstein, P., Naus, M.J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development, 46,* 818-830.

Ozonoff, S. (personal communication).

Phillipot, P., & Feldman, R.S. (1990). Age and social competence in preschooler's decoding of facial expressions. *British Journal of Social Psychology, 29,* 43-54.

Scholl, B.J. & Tremoulet, P.D. (2000). Perceptual causality and animacy. *Trends in Cognitive Science, 8,* 299-309.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64,* (Serial No. 256).

Springer, K., Meier, J.A., & Berry, D. (1996). Nonverbal bases of social perception: developmental change in sensitivity to patterns of motion that reveal interpersonal events. *Journal of Nonverbal Behavior, 20,* 199-211.

Stone, V.E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropschologia, 41,* 209-220.

Strayer, J. (1980). A naturalistic study of empathic behaviors and their relation to affective states and perspective-taking skills in preschool children. *Child Development, 51,* 815-822.

Taylor, M., Lussier, G.L., & Maring, B.L. (2003). The distinction between lying and pretending. *Journal of Cognition and Development, 4,* 299-323.

Uller, C., & Nicholls, S. (2000). Goal attribution in chimpanzees. *Cognition, 76,* B27-B34.

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis about theory of mind development: The truth about false belief. *Child Development, 72,* 655-684.

Wellman, H.M. & Hickling, A.K. (1994). The mind's "I"" Children's conceptions of the mind as an active agent. *Child Development, 65,* 1564-1580.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9,* 257-270.

Williams, K.T., & Wang, J.J. (1997). Technical references to the Peabody Picture Vocabulary Test-Third Edition (PPVT-III). Circle Pines, MN: American Guidance Service.

Table 1: Relationship of Social Cognition Measures

Measures	FP Task	MS Task (W/ faces)	MS Task (W/O faces)	Animations ToM (intent.)	Animations ToM (approp.)	Animations GD (intent.)	Animations GD (approp.)	Animations RD (intent.)	Animations RD (approp.)
FP Task	*								
MS Task (With faces)	.396**	*							
MS Task (W/o faces)	.285*	**.570*** **	*						
Animations ToM (intent.)	.227	**.364** **	**.391** **	*					
Animations ToM (approp.)	.382**	**.358** **	**.425** **	**.757*** **	*				
Animations GD (intent.)	.077	**.345** **	**.412** **	**.434** **	**.304***	*			
Animations GD (approp.)	-.023	**.177**	**.278***	**.191**	**.153**	**.465*** **	*		
Animations RD (intent.)	-.030	**.377** **	**.147**	**.313***	**.162**	**.338**	**.124**	*	
Animations RD (approp.)	.101	**-.256***	**.027**	**-.317***	**-.126**	**-.205**	**.025**	**-.818*** **	*

***p<.05**
****=p<.01**
***=p<.001

Table 2: Subject Characteristics

	N	**Mean age (months)**	**sd**	**Age range**	**Mean VMA (months)**	**VMA sd**	**VMA range**
Females	24	111.67	20.58	73-143	153.67	38.50	93-222
Males	37	110.89	17.99	74-148	140.70	27.32	96-211

Table 3: Ranking of Faux Pas Items by Proportion of Children Responding Correctly

Item Rank	**Proportion Correct**
1	.37
2	.50
3	.57
4	.58
5	.62
6	.68
7*	.78
8*	.80
9*	.83
10*	.93

* indicates items contained negatively valenced word or phrase

Table 4: Mean Score on Movie Stills Task by Emotion

Condition	**Emotion**					
	happy	sad	afraid	angry	surprised	neutral
With Faces	.99	.55	.59	.49	.50	.36
Without Faces	.71	.46	.45	.46	.39	.31

Table 5. Animations Scores

Score Type	**Condition**							
	ToM mean	ToM sd	GD mean	GD sd	RD mean	RD sd	F	p
Intentionality	3.51	1.05	2.26	.43	.67	.72	26.46	<.001
Appropriateness	.84	.38	1.40	.41	1.43	.56	17.65	<.001
Length	4.62	2.35	3.06	1.91	2.42	1.18	15.59	<.001

Table 6. Descriptive Statistics

	N	Mean	SD	Min.	Max.
Age	61	111.5	18.7	73	148
VMA	61	145.8	32.51	93	145.8
FP total	60	34.02	3.43	22	40
MS (w/o faces)	61	.48	.10	.20	.76
MS (w/faces)	61	.59	.10	.31	.85
MS (difference)	61	.11	.10	-.15	.35
RD intentionality	61	.67	.72	.00	3.17
GD intentionality	61	2.26	.43	1.17	3.33
ToM intentionality	60	3.51	1.05	1.33	6.00
RD appropriateness	61	1.43	.56	.00	2.00
GD appropriateness	61	1.40	.41	.00	2.00
ToM appropriateness	60	.84	.38	.00	1.88
RD length	61	2.42	1.18	.33	6.33
GD length	61	3.06	1.91	.00	10.5
ToM length	60	4.62	2.35	.25	12.38

Table 7. Faux Pas Correlations

	Age	**VMA**	**FP total score**
Age	*		
VMA	.648**	*	
FP total score	.296*	.302*	*

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 8. Movie Stills Correlations

	Age	**VMA**	**MS (w/o faces)**	**MS (w/faces)**	**MS (difference)**
Age	*				
VMA	.648**	*			
MS (w/o faces)	.483**	.427**	*		
MS (w/faces)	.432**	.345**	.570(**)	*	
MS (difference)	-.072	-.103	-.494(**)	.433(**)	*

Figure 1. Faux Pas Item Level of Difficulty by Age

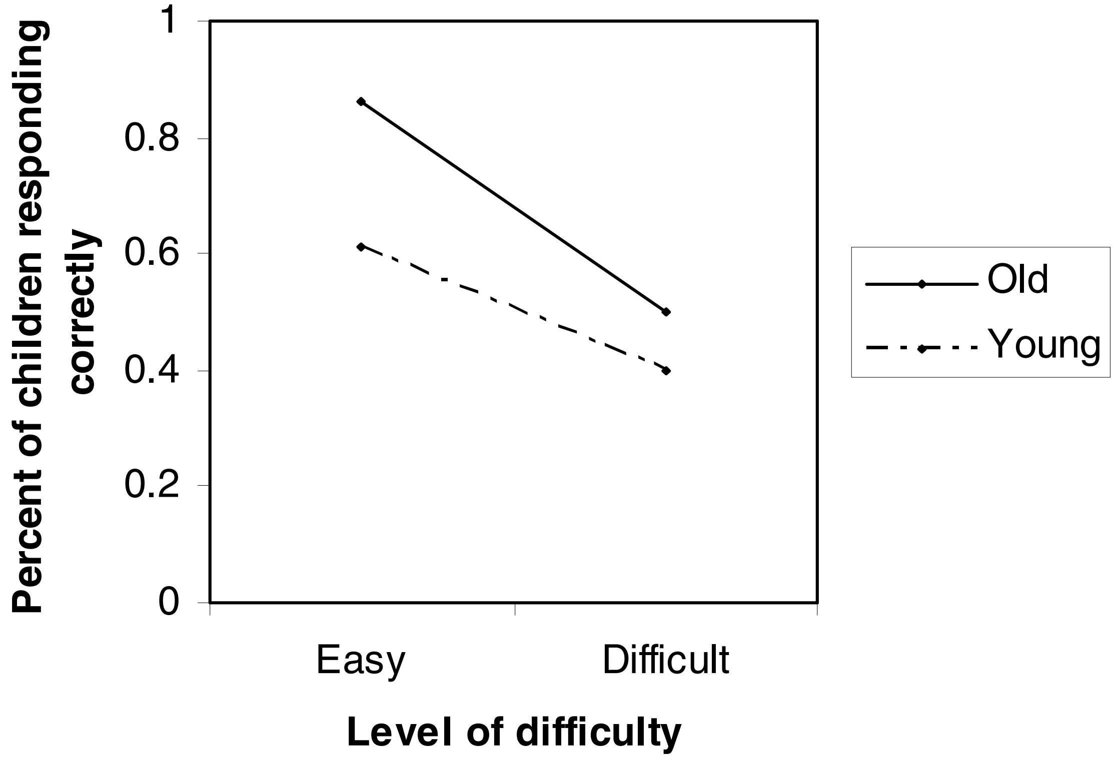

Figure 2. Random Animations Age by Gender Interaction