ABSTRACT: Purpose: A hallmark of autistic spectrum disorder (ASD) is disruption in theory-of-mind development, including the understanding of false beliefs. Previous studies have typically assessed the development of first- and second-order false belief concepts in ASD, with tasks primarily emphasizing physical causality and logical inferencing. The present study investigated how preadolescents and adolescents with ASD performed on false belief tasks that included social inferencing of psychological states as well as logical inferencing of physical states. Method: Two categories of tasks were administered: four traditional logical inferencing tasks and four social inferencing tasks specifically developed for this study. In addition, a prompt hierarchy was included to ascertain if performance on both task types improved. Participants were 45 children and adolescents primarily selected from three urban school districts: 15 adolescents with a previous diagnosis of ASD (ASD group); 15 typically developing children matched for age, gender, and ethnicity (CA group); and 15 typically developing children matched for language age, gender, and ethnicity (LA group). Results: Three findings were pertinent. First, the CA group performed at higher levels than did the LA group and the ASD group on both task sets. Second, although the CA and the LA groups performed equally well on both the logical and the social inferencing tasks, the ASD group performed better on the social inferencing tasks. Finally, the prompt hierarchy significantly improved overall task performance for the ASD and LA groups. Clinical Implications: These findings indicate that task type, variations in vocabulary ability, and the provision of support influenced performance on the false belief tasks.

KEY WORDS: autism spectrum disorder, theory of mind, false belief tasks, logical inferencing, social inferencing
A brief overview of the basic development of false belief understanding is presented first. This is followed by a description of the standard tasks used to assess the first- and second-order development of false beliefs.

**Overview of Typical Development**

To develop the basic belief–desire reasoning that supports their intentional actions, children must understand the nature and causes of emotions (Wellman & Banerjee, 1991; Wellman & Bartsch, 1988; Wellman & Hickling, 1994). To say that Ryan wanted a macadamia nut ice cream cone and remembered he could buy it at the Cheese Factory Bakery means that: “to do something intentionally is to have a desire and to engage in the act because you believe it will help satisfy your desire” (Wellman, 1990, p. 101). Children as young as 2 1/2 years can infer situationally based or perceptually displayed (objective) emotion (Wellman, Phillips, & Rodriguez, 2000). If a character in a story gets something she wants (desires), such as a macadamia nut ice cream cone, then that situation leads to a perception of happiness; if the character does not get what she wants, the outcome may be a perception of sadness. At the same time, many 2 1/2-year-olds now have schemas that allow them to understand emotions as subjective feelings that can be independent of specific situations or facial expressions; for example, getting a macadamia nut ice cream cone does not always make them happy (Wellman et al., 2000). This capacity to employ schemas that increasingly transcend particular experiences, or to go beyond the information provided in a specific situation to a more general understanding, is the basis of inferencing (Duchan, in press).

However, a more fully developed theory of mind requires the coordination and comprehension of more complex subjective states, which serve as the motivations for behavior (Westby, 2002). For example, a complex emotion might be the simultaneous experiencing of guilt and pride at having beaten out one’s best friend in a spelling contest, and these conflicting emotions require a resolution. This advanced development, which continues into the senior years (Happé, Winner, & Brownell, 1998), also allows children to share in and empathize with the social dilemmas that others’ emotional states often pose (Donahue & Foster, in press). Thus, from a reality perspective, cognitive, social–emotional, and communicative development are interdependent (Barnett & Ratner, 1997), although in practice, these domains are often assessed as separate and distinct from one another.

**Assessing False Beliefs: The Standard Tasks**

As noted earlier, children’s development of beliefs includes their understanding of false beliefs (Wellman, 1990). The standard approach to assessment involves first- and second-order belief tasks. It should be noted that most theory-of-mind studies have been conducted with typically developing preschool-age children (Miller, 2000) from primarily middle class backgrounds.

**First-order beliefs.** The earliest developing beliefs are first-order beliefs, which are generally consolidated by ages...
3 to 4½ years in the middle class samples studied (Bartsch & Wellman, 1995; Wellman & Bartsch, 1988). These first-order beliefs represent the developmental phases through which children come to understand that, although beliefs can be distinct from desires, beliefs need to be connected to desires (Wellman, 1990). In effect, the same desire (e.g., need, want, wish) can lead to different actions depending on the actor’s motivation. These early beliefs generally emerge in an ordered sequence of (a) inferred belief, (b) not-own belief, and (c) explicit false belief (Bartsch & Wellman, 1995; Wellman, 1990; see Appendix A). All share object location as the focus of predicting what another will do.

**Inferred beliefs** are not false beliefs because there is no conflict between desire and motivation; for example (modified from Wellman & Bartsch, 1988):

This morning Carlos saw his glasses on the shelf, not on the table. Now he wants his glasses. Where will he look for them?

Here, the child must predict the outcome of where Carlos will look for his glasses based on the fact that Carlos saw his glasses on the shelf and he now wants them (the character’s desire, but not belief, is explicitly stated). The correct answer to the inferred belief task goes beyond a simple location inference because the child needs some understanding of physical causality to infer the correct belief (object location) that guides Carlos where to look.

The next type, **not-own belief**, is a transition between “true” and false beliefs (Wellman, 1990). Again, only the character’s desire is overtly mentioned. To control for the child relying on his or her personal experience about where an object most likely might be located, an opposite belief is presented, which then becomes the basis for prediction (modified from Wellman & Bartsch, 1988):

This is Pam. She wants to find her puppy. It might be hiding in the house or outside. Where do you think Pam’s puppy is hiding? [Child responds with a location] That’s a good guess. Pam thinks her puppy is ______ [The opposite location is given.] Where will Pam look for her puppy?

A conflict now exists between the child’s belief and the character’s belief. To respond appropriately, the child’s own belief must be disregarded.

With the **explicit false belief** task, both the character’s desire and his or her belief (as signaled by the cognitive verb think) are specifically stated; however, the character’s belief is false relative to object location; for example (Sparrevohn & Howie, 1995, p. 255):

This is Mary. Mary wants to find her kitten. Mary’s kitten is really in the bedroom. Mary thinks her kitten is in the kitchen. Where will Mary look for her kitten? Where is it really?

To derive the correct inference about object location, young children must be capable of considering two alternate beliefs simultaneously—their own and the conflicting belief of the character. As Wellman (1990) noted, the ability to reconcile that “different people can have different beliefs (and this accounts for why they engage in different actions)” (p. 78), is a critical achievement.

**Second-order beliefs**. These more sophisticated beliefs emerge in the late preschool and early elementary age years. Children now begin to think more ably about their own thinking in regard to another’s thoughts and beliefs, as well as what they think that others think about their thoughts (Perner & Wimmer, 1985). The understanding of second-order (false) beliefs, particularly the implicit false belief, is often considered a “litmus test” (Bauminger & Kasari, 1999, p. 81) of a theory of mind because of the complexity of inferencing; however, a more conservative interpretation is that this type of reasoning serves as an index of general belief reasoning (Wellman, 1990). The child must coordinate multiple perspectives about what two characters are thinking in the situation where the conflicting beliefs of the two characters are known to the child but not to the characters (see Appendix A for an example).

**An Alternate View of Task Construction**

The standard first- and second-order false belief tasks are concerned with desires and beliefs related to object displacement. In other words, in these standard tasks, to predict the characters’ actions based on their beliefs, which are not congruent with what the child knows or believes, the child must reason about what the characters believe in relation to the movement of physical referents and their possible changes in location (Symons, McLaughlin, Moore, & Morine, 1997). Therefore, the conclusions the child derives are more or less grounded to factual reasoning about an existing reality that is comprised by physical causality. This kind of reasoning is more characteristic of logical inferencing, which also draws on the language of objective reality (Tager-Flusberg, 1993) to “talk about” what characters believe and where they will or will not search to find a dislocated object.

Although social reasoning is implied in the standard false belief tasks, for example, a father wants to “surprise” a daughter with a birthday present, these tasks appear less often to elicit inferences about feelings and attitudes that distinguish interpersonal relationships (Johnson & von Hoff Johnson, 1986). This type of social reasoning is more often grounded to pragmatic inferences than to logical inferences. Pragmatic inferences draw on individuals’ world (or situational) knowledge of relations between people and events. Most importantly, pragmatic inferencing is related to concepts of psychological causality more so than to concepts of physical causality (Westby, 2002). In pragmatic inferencing, children must center their attention on the conflicting social beliefs and desires of characters while also paying attention to the relevant situational factors (Budwig, 1999; Symons et al., 1997), as this scenario illustrates (Baron-Cohen, O’Riordan, Stone, Jones, & Piaisted, 1999, p. 416):

Mrs. West, the teacher, had something to tell her class, “one of the boys in our class, Simon, is very seriously
ill” she said. The class were all very sad and were sitting quietly when a little girl, Becky arrived late. “Have you heard my new joke about sick people?” she asked. The teacher said to her “Sit down and get on with your work.” What did the teacher tell the class at the beginning of the story? Did Becky know Simon was sick?

If asked whether Becky would be embarrassed and why she might feel this way, there are plausible alternatives that might be considered about the sources of Becky’s mistaken belief and its consequences. For example, it is most likely that she did not know that Simon was sick and, therefore, one might predict that she would now feel mortified about her poor taste in jokes. This kind of pragmatic inferencing may also rely more heavily on the language of mental states (Westby, 2002), such as “Becky said that she was sorry about her behavior,” rather than depend on the language of objective reality (Tager-Flusberg, 1993), which expresses “facts” of reality, such as “Sit down and get on with your work.”

A significant point is that individual differences are apparent in theory-of-mind development, as assessed through the standard tasks. These variations, which are not understood well, are attributed to task variables (Miller, 2000; Wellman et al., 2001), sociocultural variations in family socialization practices and socioeconomic status (Astington, 1998a; Bartsch & Estes, 1996; Cutting & Dunn, 1999; Lillard, 1998; Ozonoff, 2001; Vinden, 1999), and differences in language proficiency (e.g., Astington, 1998b, 2001; Astington & Jenkins, 1999; de Villiers & de Villiers, 2000). For example, Astington (2001) suggested that earlier developing concepts of desire and intention might be less dependent on language ability, whereas false belief understanding may co-vary more with difference in language skills. de Villiers and de Villiers speculated that, because sentences containing mental state verbs often require complements, children must have the syntactic ability to engage in false complement processing, such as the processing of “Dad thought that he forgot the birthday present,” before they can provide an adequate mental explanation for contradictory perspectives. Wellman et al. (2001) argued otherwise; that is, changes in conceptual knowledge drive the language of beliefs and false beliefs and, further, the medium of presentation, such as the use of a real person versus pictured storyboard characters, or question type are not significant variables influencing performance on false belief tasks.

Regardless of these unresolved controversies, it is clear that a typically developing theory of mind consists of multiple components, including language components. From a developmental perspective, the cognitive and social components of a theory of mind must become integrated in order for children to make complex inferences about the content of others’ minds (Tager-Flusberg, 2001). Among these contents are integrated concepts of physical and psychological causality, both of which contribute to and support different aspects of false belief reasoning. However, the standard false belief tasks employed in ASD have primarily emphasized physical causality, not psychological causality.

**FALSE BELIEF PERFORMANCE IN ASD ON THE STANDARD TASKS**

A general agreement is that disruptions in false belief development are not unique to ASD, although the severity of delay may be greater than in other child populations, such as those with mental retardation, schizophrenia, or language impairment (Bishop, 2000; Happé, 1995; Pilowsky, Yirmiya, Arbelle, & Mozes, 2000; Serra et al., 2002; Tager-Flusberg, 2001, 2003; Ziatas, Durkin, & Pratt, 1998). In view of this consensus that breakdowns in false belief reasoning are not confined to ASD, an important question concerns what the standard second-order false belief task tells about social inferencing ability in ASD.

**Performance Patterns on Second-Order False Belief Tasks**

In preadolescents, adolescents, and young adults with ASD, including those with Asperger’s syndrome, a reliable finding has been disrupted performance on the standard false belief tasks (Baron-Cohen, Leslie, & Frith, 1985; Bowler, 1992; Happé, 1995; Leslie & Frith, 1988; Ozonoff, Pennington, & Rogers, 1991; Perner, Frith, Leslie, & Leekam, 1989; Sparrevohn & Howie, 1995; Tager-Flusberg & Sullivan, 1994a). Results also consistently show that, similar to typically developing but much younger children, first-order beliefs are understood before second-order beliefs.

But there are contrary patterns. Two important findings are that, first, some samples with ASD who are higher functioning can pass a second-order belief task, even as young as age 6;11 (years;months; Ziatas et al., 1998), whereas others who pass are at least in their early or mid-adolescence. Second, the pass rate is typically correlated with higher verbal ability as measured by standardized vocabulary tests (e.g., Happé, 1995; Kazak, Collis, & Lewis, 1997; Sparrevohn & Howie, 1995; Tager-Flusberg & Sullivan, 1994b; Ziatas et al., 1998). These findings indicate that differences in word learning proficiency might matter. Some higher functioning individuals with ASD can pass standard second-order tasks provided they reach a point in their general vocabulary size that seems associated with successful performance. A qualification is that passing the second-order belief question does not mean that participants are also able to justify their answer appropriately (Bauminger & Kasari, 1999), for example, explaining the reason for their response. As some researchers have found (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen et al., 1993; Baron-Cohen & Swettenham, 1997; Happé, 1994; Heavey, Phillips, Baron-Cohen, & Rutter, 2000; Hillier & Allinson, 2002), the fact that a minority of older and higher functioning individuals with ASD pass a second-order false belief task cannot be interpreted as indicating an intact theory of mind because they still have significant problems with understanding more complex emotion and resolving real-life social quandaries. Nevertheless, studies showing that some older individuals with ASD can pass the second-order belief task
have all used the standard task in which physical causality and more logical inferencing dominate in contrast with psychological causality and social inferencing; thus, previous work sheds little light on how the social inferencing skills of adolescents with ASD may be accessible on first- or second-order false belief tasks.

**Independent Versus Assisted Performance**

Regardless of the participants’ ages, the vast majority of studies assessing first- or second-order beliefs employed a pass–fail scoring system; responses were scored as either correct or incorrect (e.g., Baron-Cohen et al., 1985; Bauminger & Kasari, 1999; Bowler, 1992; Happé, 1994, 1995; Muris et al., 1999; Pernar et al., 1989; Pilowsky et al., 2000; Sparrevohn & Howie, 1995; Ziatas et al., 1998). This kind of elicitation and scoring format reflects two mistaken notions. One is that “theory of mind is something one does or does not have” (Tager-Flusberg, 2001, p. 177) at a given point in time; the other assumes that assessing a domain of knowledge, such as false beliefs, as an “independent performance” presents a complete picture of individual competence. In other words, a presumption of these studies appeared to be that, if knowledge is already developed in some manner, it can be accessed and applied without assistance (Palincsar, Brown, & Campione, 1994).

One challenge to the notion of independent performance is that overall performance on false belief tasks may improve with some form of guided assistance, such as prompting. This type of support may promote greater ease in retrieval of knowledge not fully developed because it enhances the prominence of the specific information to be recalled (Rieffe, Terwogt, & Stockmann, 2000). Systematic prompting is also consistent with the general notion of scaffolding (Stone, 2002), in which children’s potential to learn is discerned from changes in performance when varying levels of guided assistance are provided. Research by Ozonoff and Miller (1995) suggested that certain theory-of-mind concepts are “learnable” when appropriate supports are provided (e.g., teaching how to think about theory-of-mind activities before false belief tasks are administered). However, the actual (or unassisted) performance of adolescents with autism on theory-of-mind tasks is the only data serving as the bases for current conclusions.

In summary, three points are pertinent.

- Regardless of the theoretical framework applied to account for disrupted development, impairments in theory of mind, while not exclusive to ASD, are considered a major feature of the significant socio-cognitive and communicative difficulties confronting individuals with ASD.
- Current studies on false belief development in ASD have primarily relied on the standard tasks, which appear to stress physical causality and, hence, logical inferencing, and depend as well on the language of action that encodes “objective reality” concepts (Tager-Flusberg, 1993). The standard tasks have yet to incorporate psychological causality where social inferencing is primary for understanding the concept that other people’s feelings or attitudes, as well as their physical actions, are caused by their beliefs or, even, their false beliefs. This omission is striking given the fact that the social world of the mind is inherent to a basic definition of a theory of mind.
- Studies have investigated performance on theory-of-mind tasks through the prism of participants’ independent performance. The role of guided assistance in the form of varying levels of support to enhance performance remains to be explored.

The purpose of this study was twofold. The first intent was to construct theory-of-mind tasks that preserved the same format as the logical inferencing tasks used in previous studies with adolescents while altering the focus of these tasks. A specific aim was to probe the ability of adolescents with and without autism to infer cognitive versus social mental states comprising first- and second-order beliefs tasks. The second purpose was to incorporate a graduated prompt hierarchy into task construction. The objective was to determine individual potential for emerging concepts of false beliefs in adolescents with ASD in contrast to a chronological age-matched group and a language age-matched group.

Three questions were addressed. The first question asked what effect ASD had on the child’s ability to perform on false belief tasks. A second question addressed whether the type of inferencing requirement (i.e., logical vs. social inferencing) differentially affected performance for the three groups. The final question concerned whether a graduated prompt hierarchy affected performance on the logical and social inferencing tasks for the three groups as a function of inferencing type.

**METHOD**

**Participants**

A total of 45 participants, with 15 participants in each of three groups, served as the final sample. These groups were a group with autism spectrum disorder (ASD), a chronological age-matched (CA) group, and a language age-matched (LA) group. Participants were individually matched for group membership by gender, ethnicity, and either chronological or language age. Each group consisted of 13 males and 2 females as well as 13 Caucasian and 2 African American participants.

Participants with ASD were recruited from two urban school districts and a private speech-language pathology practice primarily serving children and adolescents with ASD. Inclusion in the final sample was based on (a) school records containing medical documentation of ASD in the form of autistic disorder or Asperger’s syndrome and (b) individual ability to participate in the administration of the Peabody Picture Vocabulary Test–III (PPVT–III; Dunn & Dunn, 1997), a vocabulary recognition measure. The LA group was gathered from two urban school districts and word-of-mouth advertising. Participants in this group were individually matched within the PPVT–III age-equivalent
range to each participant in the ASD group. The CA group was gathered from two urban school districts.

As seen in Table 1, the mean chronological ages for the CA group and the ASD group, as well as the mean age equivalents on the PPVT–III between the LA group and the ASD group, were within 1 month of each other. Mean standard scores on the PPVT–III for both the CA and the LA groups were within or above the average range, with the exception of two participants, one in the CA group and one in the LA group. The mean standard scores of the ASD group were ~2 SD below the mean of 100 (see Table 2 for individual participant information).

In addition to the formal vocabulary measure, the classroom teachers of participants in the ASD group completed a questionnaire designed to obtain information about the typical school day. These questionnaires indicated that 11 of the 15 students reportedly spent 100% of each day in the special education classroom. Three students participated in regular education classes for an average of 58% of each day, and 1 student was fully included in regular education classes with the assistance of a full-time paraprofessional.

Task Construction

The theory-of-mind tasks were divided into two sets—logical inferencing and social inferencing. The logical inferencing tasks were the standard tasks that highlighted physical causality; the social inferencing tasks were developed for this study and emphasized psychological causality.

Four narrative scenarios were constructed for each set. The five logical inferencing scenarios were modified from investigations with typically developing children, ages 2–10 years (Wellman & Bartsch, 1988), children with autism (Baron-Cohen et al., 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994), and children who were deaf (Giaccone, 1997). Parallel scenarios were then created that focused on social inferencing and corresponded with the developmental sequence previously established for the logical inferencing tasks. This developmental sequence consisted of four levels: inferred belief, not-own belief, explicit false belief, and implicit false belief.

Logical inferencing tasks. Three first-order tasks (1 to 3) and one second-order task (4) were developed, as shown in Appendix A. These first-order logical inferencing tasks were modified from the Wellman and Bartsch (1988) tasks.

1. Inferred belief: In this first-level task, a character’s desire is explicitly stated, but the character’s belief is not (Wellman, 1990). Participants are asked to infer the character’s belief based on the events depicted in the story and then predict the character’s action in accord with that inferred belief.

2. Not-own belief: The second concept in first-order development involves a task in which the character in the narrative holds an opposite belief from the participants, who does not know the actual location of the missing object. To predict the character’s action, participants must put aside their own beliefs and acknowledge the character’s opposing belief. Figure 1 provides an illustration of a script and the pictures used for this task.

3. Explicit false belief: In this more advanced phase of development, participants are expected to predict what a character will do based on knowledge of the character’s false belief. To predict the character’s action, participants need to suppress what they know to be true and acknowledge the character’s explicitly stated false belief.

4. Implicit false belief: Achievement of a second-order concept was assessed by an implicit false belief task, as modified from Sullivan et al. (1994). In this logical inferencing task, the character’s false belief is not explicitly stated. The character’s response must be predicted based on inferences about that character’s false belief about another character’s belief.

Social inferencing tasks. To examine the ability to infer psychological states motivated by desires and beliefs, four social inferencing tasks (as shown in Appendix B) were

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean chronological age</th>
<th>Mean age equivalent (PPVT–III)</th>
<th>Mean standard score (PPVT–III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>9;3 (4;2–11;11)</td>
<td>7;8 (5;8–13;1)</td>
<td>102.3 (80–115)</td>
</tr>
<tr>
<td>ASD</td>
<td>14;5 (11;7–20;0)</td>
<td>7;9 (5;8–12;0)</td>
<td>66.9 (40–101)</td>
</tr>
<tr>
<td>CA</td>
<td>14;4 (11;6–19;5)</td>
<td>16;3 (9;7–22;0)</td>
<td>103.0 (86–130)</td>
</tr>
</tbody>
</table>

Note. LA = language age-matched group, ASD = group with autism spectrum disorder, CA = chronological age-matched group.

---

It is recognized that a language age value, similar to a mental age value, is a psychometric unit that derives from a group’s average performance only and cannot be interpreted as indicating equivalent performance. However, verbal (mental) age (VMA) is widely used in many studies on ASD (Bauminger & Kasari, 1999; Happé, 1995; Ozonoff & Miller, 1995; Sparrevohn & Howie, 1995). Moreover, the obtained VMA was often based on the same vocabulary measure used in this study, although determined from earlier versions. The concept was assessed by an implicit false belief task, as modified from Sullivan et al. (1994). In this logical inferencing task, the character’s false belief is not explicitly stated. The character’s response must be predicted based on inferences about that character’s false belief about another character’s belief.
Table 2. Participant information for the three groups (N = 45) and descriptive data on the PPVT–III and the logical and social inferencing tasks.

<table>
<thead>
<tr>
<th>Age (in years;months)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>PPVT–III standard score</th>
<th>PPVT–III age equivalent</th>
<th>Logical 1st-order tasks (n = 3)</th>
<th>Logical 2nd-order task (n = 1)</th>
<th>Social 1st-order tasks (n = 3)</th>
<th>Social 2nd-order task (n = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism Spectrum Disorder Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11;7</td>
<td>F</td>
<td>C</td>
<td>75</td>
<td>7;2</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>11;9</td>
<td>M</td>
<td>C</td>
<td>101</td>
<td>12;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;1</td>
<td>M</td>
<td>C</td>
<td>58</td>
<td>5;8</td>
<td>0/3</td>
<td>0/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>12;5</td>
<td>M</td>
<td>AA</td>
<td>77</td>
<td>8;3</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>12;7</td>
<td>M</td>
<td>C</td>
<td>78</td>
<td>8;6</td>
<td>1/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>12;8</td>
<td>M</td>
<td>C</td>
<td>60</td>
<td>6;0</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;10</td>
<td>M</td>
<td>C</td>
<td>78</td>
<td>8;8</td>
<td>1/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>14;2</td>
<td>M</td>
<td>C</td>
<td>52</td>
<td>6;4</td>
<td>0/3</td>
<td>0/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>14;10</td>
<td>M</td>
<td>C</td>
<td>74</td>
<td>9;0</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>15;1</td>
<td>M</td>
<td>C</td>
<td>47</td>
<td>6;4</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>15;2</td>
<td>M</td>
<td>AA</td>
<td>76</td>
<td>9;7</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>16;1</td>
<td>M</td>
<td>AA</td>
<td>40</td>
<td>Up to 6;0</td>
<td>3/3</td>
<td>0/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>16;1</td>
<td>M</td>
<td>C</td>
<td>76</td>
<td>9;7</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>19;5</td>
<td>F</td>
<td>C</td>
<td>40</td>
<td>Up to 6;6</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>20;0</td>
<td>M</td>
<td>C</td>
<td>72</td>
<td>9;6</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
</tbody>
</table>

Language Age-Matched Group

<table>
<thead>
<tr>
<th>Age (in years;months)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>PPVT–III standard score</th>
<th>PPVT–III age equivalent</th>
<th>Logical 1st-order tasks (n = 3)</th>
<th>Logical 2nd-order task (n = 1)</th>
<th>Social 1st-order tasks (n = 3)</th>
<th>Social 2nd-order task (n = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4;2</td>
<td>M</td>
<td>C</td>
<td>130</td>
<td>6;9</td>
<td>2/3</td>
<td>1/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>5;7</td>
<td>M</td>
<td>C</td>
<td>106</td>
<td>6;1</td>
<td>2/3</td>
<td>0/1</td>
<td>2/3</td>
<td>1/1</td>
</tr>
<tr>
<td>5;8</td>
<td>M</td>
<td>C</td>
<td>108</td>
<td>6;6</td>
<td>2/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>6;1</td>
<td>M</td>
<td>AA</td>
<td>95</td>
<td>5;8</td>
<td>2/3</td>
<td>0/1</td>
<td>1/3</td>
<td>1/1</td>
</tr>
<tr>
<td>6;2</td>
<td>F</td>
<td>C</td>
<td>106</td>
<td>6;1</td>
<td>2/3</td>
<td>0/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>6;7</td>
<td>F</td>
<td>C</td>
<td>93</td>
<td>6;0</td>
<td>2/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>6;10</td>
<td>M</td>
<td>C</td>
<td>80</td>
<td>6;0</td>
<td>2/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>7;1</td>
<td>M</td>
<td>C</td>
<td>115</td>
<td>8;6</td>
<td>2/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>7;10</td>
<td>M</td>
<td>C</td>
<td>102</td>
<td>8;2</td>
<td>2/3</td>
<td>0/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>8;6</td>
<td>M</td>
<td>AA</td>
<td>95</td>
<td>8;2</td>
<td>2/3</td>
<td>1/1</td>
<td>1/3</td>
<td>1/1</td>
</tr>
<tr>
<td>8;11</td>
<td>M</td>
<td>C</td>
<td>100</td>
<td>8;1</td>
<td>2/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>9;0</td>
<td>M</td>
<td>C</td>
<td>102</td>
<td>9;6</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>9;6</td>
<td>M</td>
<td>C</td>
<td>104</td>
<td>10;3</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>0/1</td>
</tr>
<tr>
<td>10;4</td>
<td>M</td>
<td>C</td>
<td>93</td>
<td>9;4</td>
<td>3/3</td>
<td>1/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
<tr>
<td>11;11</td>
<td>M</td>
<td>C</td>
<td>105</td>
<td>15;0</td>
<td>2/3</td>
<td>1/1</td>
<td>2/3</td>
<td>0/1</td>
</tr>
</tbody>
</table>

Chronological Age-Matched Group

<table>
<thead>
<tr>
<th>Age (in years;months)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>PPVT–III standard score</th>
<th>PPVT–III age equivalent</th>
<th>Logical 1st-order tasks (n = 3)</th>
<th>Logical 2nd-order task (n = 1)</th>
<th>Social 1st-order tasks (n = 3)</th>
<th>Social 2nd-order task (n = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11;6</td>
<td>F</td>
<td>C</td>
<td>114</td>
<td>15;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>11;10</td>
<td>M</td>
<td>C</td>
<td>102</td>
<td>12;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;2</td>
<td>M</td>
<td>C</td>
<td>89</td>
<td>10;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;11</td>
<td>M</td>
<td>AA</td>
<td>86</td>
<td>10;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;8</td>
<td>M</td>
<td>C</td>
<td>106</td>
<td>14;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>12;9</td>
<td>M</td>
<td>C</td>
<td>115</td>
<td>17;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>13;2</td>
<td>M</td>
<td>C</td>
<td>106</td>
<td>14;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>14;2</td>
<td>M</td>
<td>C</td>
<td>78</td>
<td>9;07</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>14;6</td>
<td>M</td>
<td>C</td>
<td>130</td>
<td>22;0+</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>14;9</td>
<td>M</td>
<td>C</td>
<td>92</td>
<td>12;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>15;8</td>
<td>M</td>
<td>C</td>
<td>116</td>
<td>22;0+</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>16;0</td>
<td>M</td>
<td>AA</td>
<td>96</td>
<td>14;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>15;6</td>
<td>M</td>
<td>C</td>
<td>110</td>
<td>19;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>19;5</td>
<td>F</td>
<td>C</td>
<td>97</td>
<td>17;0</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>19;4</td>
<td>M</td>
<td>C</td>
<td>114</td>
<td>22;0+</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
</tr>
</tbody>
</table>

Note: PPVT–III = Peabody Picture Vocabulary Test–III (Dunn & Dunn, 1997); Gender: M = male, F = female; Ethnicity: C = Caucasian, AA = African American. For the false belief tasks, the number of tasks passed is shown for each participant regardless of whether support was needed.

devolved to parallel the four developmental levels of the logical inferencing tasks. The social inferencing tasks (5–8) required participants to predict a character’s feelings based on their understanding of the character’s beliefs and affective reactions to events depicted in the stories.

5. Inferred belief: The first social inferencing task corresponded to the standard inferred belief task: The answer must be inferred from the information given regarding the emotion the character reportedly felt.

6. Not-own belief: The second social inferencing task corresponded to the standard not-own belief task. Participants had to suspend their own belief about
how the character felt and respond according to the feeling stated in the story (see Figure 2 for the script and accompanying illustrations).

7. **Explicit false belief**: The third social inferencing task paralleled the standard explicit false belief task. This task required participants to infer how a character would feel upon realizing that the character’s belief was false.

8. **Implicit false belief**: The last social inferencing task was a second-order task that required the multiple inferential integration of “Dad thinks that Frank feels...”.

**Materials and Instrumentation**

Figures 1 and 2 display illustrations that accompanied the logical and social inferencing task sets (ClickArt® Incredible Image Pak 65,000, 1996; Corel® Mega Gallery, 1996). A binder containing the illustrations for each set of tasks was used to present the scenarios.

All testing sessions were audio-recorded.

**Procedure**

Each participant was tested in one session on the logical inferencing tasks (tasks 1–4) and the social inferencing tasks (tasks 5–8). The presentation of task set was counterbalanced across the participants within each group. All testing was administered by one of five examiners, all of whom were graduate students in speech-language pathology. To minimize examiner variations in intonational contours, the scripts were read emphasizing key words with two intonational patterns, derived from Levelt (1989), notated on the scripts. These contours were high fall tone (falling intonational contour), associated with declarative statements, and high rise tone, which signaled a question.

**Administration.** Before administration began, the examiner read a brief set of instructions, including a simplified purpose of the tasks. Participants were told that they were to help the examiner decide whether or not little children would like the stories and why. This purpose was included because research indicates that more complex narratives result when children are provided with a shared purpose for the activity (Bamberg & Reilly, 1996). After describing the purpose to the participants, the examiner read the scenario script and the experimental questions for each task as the participant looked at the corresponding pictures. The types of questions varied with the task level (see Appendices A and B). For example, questions for the first three tasks in each set primarily elicited a prediction. Questions for the fourth task in each set probed for the understanding of false belief versus
realities and the ability to retain information in working memory.

Participants were allowed to request any part of the scenarios to be repeated once. Second, they could request clarification of any ambiguous words. Finally, in the event that a participant was not able to answer the questions following the scenarios, a graduated hierarchy of prompts, from low to high levels of support, was offered (Schneider & Watkins, 1996; Stone, 2002; see Appendices A and B). A low level of assistance meant that no additional prompts were needed for the participant to answer the question appropriately. If the participant did not answer the question, a medium level of assistance was provided through a cloze procedure, such as, “Joe will look ________.” The highest level of assistance offered before a response was coded as incorrect was a forced choice. The participant was given two alternative answers, such as, “Joe will look on the shelf or Joe will look on the table” and asked to select one or the other. This prompt hierarchy, developed specifically for the study, was implemented to reduce complexities in retrieval or with language formulation that might be competing with access of the underlying concept.

**Scoring.** During testing, the examiner recorded participants’ responses to the questions on a data form. One of four scoring categories was subsequently applied: (a) correct with no help, (b) correct with the addition of the cloze procedure, (c) correct with the presentation of the forced choice prompt, and (d) incorrect response even with graduated support. Performance was measured as number of task level questions correctly answered. Participants who answered correctly in the no help condition were automatically credited with the cloze and forced choice options. The participant was given two alternative answers, such as, “Joe will look on the shelf or Joe will look on the table” and asked to select one or the other. This prompt hierarchy, developed specifically for the study, was implemented to reduce complexities in retrieval or with language formulation that might be competing with access of the underlying concept.

**Examiner Agreement**

Five raters, all graduate students in speech-language pathology, independently viewed a videotaped session for 30% of the total sample. Cohen’s kappa was used to determine the consistency of response coding. A kappa of 1.0 for the logical inferencing tasks and .76 for the social inferencing tasks indicated consistency among the examiners and the raters in their scoring decisions well beyond chance expectations (Maxwell & Satake, 1997).

**RESULTS**

Performance was measured as number of task level questions correctly answered. It was examined as a function of Diagnostic Group, Type of Theory-of-Mind Task, and Support condition using a $3 \times 2 \times 2$ mixed model repeated measures analysis of variance (ANOVA). Descriptive statistics for the performance of the 45 individual participants on the two sets of tasks are shown in Table 2. Table 3 displays the ANOVA results.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group (G)</td>
<td>2</td>
<td>22.146</td>
<td>.0000001*</td>
<td>.54</td>
</tr>
<tr>
<td>Task Type (T)</td>
<td>1</td>
<td>15.333</td>
<td>.000325*</td>
<td>.18</td>
</tr>
<tr>
<td>Support Level (S)</td>
<td>2</td>
<td>11.414</td>
<td>.000041*</td>
<td>.26</td>
</tr>
<tr>
<td>G x T</td>
<td>2</td>
<td>9.844</td>
<td>.000312*</td>
<td>.38</td>
</tr>
<tr>
<td>G x S</td>
<td>4</td>
<td>1.304</td>
<td>.275169</td>
<td></td>
</tr>
<tr>
<td>T x S</td>
<td>2</td>
<td>1.247</td>
<td>.294133</td>
<td></td>
</tr>
<tr>
<td>G x T x S</td>
<td>4</td>
<td>17.717</td>
<td>.0000001*</td>
<td>.38</td>
</tr>
</tbody>
</table>

*p < .001

**Main Effects**

For the purposes of data analysis, the first- and second-order tasks for each condition were combined in order to avoid comparing three tasks (the first-order tasks) with one task (the second-order task). It can be seen from Table 3 that all of the main effects were statistically significant. For the main effect of Diagnostic Group (G), mean performance equaled $2.42 (SD = 1.13)$, $2.53 (SD = .62)$, and $3.9 (SD = .26)$ for the ASD, LA, and CA groups, respectively. There was considerable variation in the performance of the ASD group. A partial Eta² statistic ($\eta^2_p = .54$) suggested that diagnostic group membership explained approximately 54% of the variance. This is a moderate effect size. Post hoc testing using the Tukey test revealed that performance for the CA group was significantly higher than for the ASD or LA groups. It is also important to note that the CA group was performing at or near ceiling levels. The difference in performance between the ASD and LA groups, however, was not statistically significant.

The means for Type of Theory-of-Mind Task (T) were $2.86 (SD = 1.19)$ and $3.04 (SD = .74)$ for logical and social inferencing, respectively. Thus, performance on the social inferencing tasks was significantly better than for the logical inferencing tasks. The partial Eta² statistic ($\eta^2_p = .18$) indicated that task type accounted for approximately 18% of the variance and is considered to be a small effect.

Finally, the respective means for the prompt Support Levels (S) were $2.82 (SD = .76)$, $2.95 (SD = .65)$, and $3.09 (SD = .60)$ for no support, no support + cloze, and no support + cloze + forced choice. Improved performance as a function of increasing support would be expected and the increase was significant. The partial Eta² statistic ($\eta^2_p = .26$) indicated that the graduated prompt levels accounted for approximately 26% of the variance (i.e., a small effect). Results of a Tukey post hoc test indicated that the means were not significantly different between the no help support condition and the support condition of no help plus the cloze procedure; but means were significantly different between the no help condition and the combined cloze procedure and forced choice levels. A significant difference was also noted between the cloze and the forced choice conditions.
Interactions

In addition to the significant main effects, a significant interaction was found between Diagnostic Group (G) and Type of Theory-of-Mind Task (T). The partial Eta² statistic ($\eta^2 = .38$) indicated that this interaction accounted for approximately 38% of the variance, which is a small-moderate effect. Performance as a function of task type and group is shown in Figure 3. As expected, it can be seen that the CA group performed better than the ASD and LA groups on both task types. Results of a Tukey post hoc test confirmed that these differences in performance were statistically significant. Tukey post hoc testing also indicated that differences in performance as a function of task type were not statistically significant for the CA or the LA group. However, the ASD group performed significantly better on the social inferencing tasks than on the logical inferencing tasks. Additionally, the ASD group performed significantly better than the LA group on the social inferencing tasks while the converse was true on the logical inferencing tasks. That is, on logical inferencing tasks, the performance of the LA group was significantly better than the performance of the ASD group.

As Table 3 further shows, neither the interaction between Diagnostic Group and Support Level nor the interaction between Task Type and Support Level was statistically significant. However, the 3-way interaction between Diagnostic Group (G), Task Type (T) and Support Level (S) reached significance. This interaction is illustrated in Figure 4. To further explore this interaction, meaningful within-group comparisons were made using paired $t$ tests and meaningful between-group comparisons were conducted using Bonferroni corrections for multiple comparisons. As expected, the CA group means were significantly higher than the means for the ASD and LA groups for both task types at all support levels. This finding was not surprising given the high level of overall performance exhibited by the CA participants.

Next, the effect of support for both the ASD and the LA groups for the logical inferencing tasks was examined. It can be seen from Figure 4 that performance for both groups increased with the use of cloze support but did not increase further when the forced choice support was added. The opposite pattern occurred for the social inferencing task for both groups. Performance did not improve with the use of cloze support but increased with the addition of forced choice support.

To summarize the pertinent findings, first, the CA group performed at higher levels than both the LA and the ASD groups on both the social and the logical inferencing tasks. Second, although the CA and LA groups performed equally well on both task types, in general, the participants with ASD performed better on the social rather than the logical inferencing tasks. Third, the prompts functioned to increase the number of participants in all groups who passed at a variety of levels on both task types.

DISCUSSION

This study addressed three interrelated questions concerning performance on first- and second-order false belief tasks that index a basic concept of the mind’s contents. The first question asked what effect ASD had on task performance. The second concerned whether the type of inferencing requirement (i.e., logical vs. social inferencing) differentially affected performance for the three groups, and the third question asked whether graduated prompts affected performance.

Effects of Group Membership

The effect of group membership was significant, with the CA group performing significantly better than the other two groups. The near ceiling performance of the CA group regardless of condition was an expected finding. An unanticipated finding was that the ASD group and the LA group performed at similar levels on overall task performance. Research indicates that, for the standard tasks, most typically developing children increasingly pass the first-order belief tasks above chance levels at age 4 years and older (Wellman et al., 2000), whereas the second-order false belief is passed with increasing accuracy after age 6 years (Wellman, 1990). Therefore, the expectation was that the LA group would perform better than the ASD group. In fact, as the individual data in Table 2 show, the only participant in the LA group to pass both task sets was an 11.6-year-old female, a finding that appears contrary to the general patterns of false belief development in typically developing children.

The similarity in profiles for overall task performance between the LA and the ASD groups indirectly supports prior studies at some level. Prior research found that general lexical knowledge, typically measured by various versions of the PPVT, was associated with false belief performance in an unspecified manner in preteens and adolescents with ASD (e.g., Happé, 1995; Kazak et al., 1997; Sparrevohn & Howie, 1995; Tager-Flusberg & Sullivan, 1994a; Ziatas et al., 1998), as well as in typically developing preadolescents (Bosacki, 2000). On the other hand, similarity in performance between the LA and the
ASD groups is not equivalent performance. Examination of individual performance patterns on task type (see Table 2) indicated that neither chronological age nor language age readily corresponded with who passed (or failed), with or without support. This pattern of marked dispersion in performance regardless of age would be generally consistent with the Wellman et al. (2001) argument that the availability of children’s concepts of beliefs about belief states, not their general lexical level, is the major variable accounting for performance differences. It is also consistent with the conclusions that, when children are still in the process of mastering false belief concepts, differences in task targets (e.g., logical vs. social inferencing tasks) will influence their reliability in attributing false beliefs accurately (Miller, 2000).

Effects of Inferencing Requirements

The robust finding that the ASD participants as a group performed more adequately on the social inferencing tasks than on the logical inferencing tasks deserves further consideration. In terms of the most stringent criterion, 33% passed all three of the first-order social inferencing tasks, as well as the second-order social inferencing task (see Table 2). In contrast, 47% passed the three first-order social inferencing tasks only. Turning to the logical inferencing tasks, 20% passed both the first- and second-order tasks, whereas 6% (n = 1) passed the first-order tasks only. This pattern suggests preliminary evidence that, with the availability of the prompt hierarchy, certain youngsters with ASD could engage in the level of social reasoning required in the social inferencing tasks to a greater degree than they could engage in the logical reasoning demanded by the standard tasks, even with support.

This result cannot be explained by findings that youngsters with ASD recognize situationally based emotions that are perceptually apparent, such as happy and sad (Baron-Cohen et al., 1993, Dennis, Lockyer, & Lazenby, 2000), or that they are less able as a group to predict more complex emotions (e.g., Baron-Cohen et al., 1999; Hillier & Allinson, 2002). Instead, performance patterns align with the limited research showing that individuals with ASD under certain circumstances can attribute more complex feelings to others (Rieffe et al., 2000), such as relief, worry, surprise, and embarrassment. All four meanings communicate the concept of empathy and were central to understanding of the first-order (explicit false belief) social inferencing scenario (relief and worry) as well as the scenario for the second-order implicit false belief task (surprise and embarrassment). The question then is why the participants with ASD, in contrast to other studies, were better able to pass these two tasks as compared with their performance on the logical inferencing tasks counterparts.

A factor possibly accounting for this pattern concerns the influence of the linguistic/discourse experiences of children with ASD. Performance on false belief tasks that require social reasoning and, hence, more pragmatic inferencing, may be enhanced or restricted based on the nature and quality of children’s participation in discourse experiences at home and at school where talk is centered on psychological causality about people's feeling states. This would include talk about children’s own feeling states. As Bamberg (2001) pointed out, when mental states are a theme of talk in real-life social contexts, learning to talk openly and effectively about others' feelings as motivations for their beliefs is wound up with how well children can construct and coordinate the speaker–audience relationship with their own self-identity. For example, social conflicts, or disputes,
with siblings create natural opportunities for children to grapple with different emotional points of view and the feelings embedded in these varied perspectives, including their own feelings (Bartsch & Estes, 1996). Moreover, it has been suggested that individuals with ASD, as they reach the adolescent years, similar to their chronological age counterparts, have increased motivation to seek out social relationships (Baron-Cohen et al., 1985; Bowl cycle, 1992). This motivation provides additional opportunities for expanding their self-identity, mental state lexicon, and social inferencing ability through a broader array of social transactions and memberships. From an educational perspective, the opportunities created through the world of literature afford access to others’ linguistic universes. Social inferencing skill is essential for “getting inside” the mental worlds of literary and historical characters, interpreting their motivations, and creating internal dialogues with the author about what the text means (Aston ton, 1998a; Donahue & Foster, in press). Thus, one possibility for explaining the better performance of the ASD group on the social inferencing tasks, even in contrast to the LA group, may be related to the overall duration of their real-world experience with social schemas, including their experiences with related pragmatic inferences, both at home and at school. More intensive investigation is warranted on the specific family and educational practices that may be effective in orchestrating and integrating children’s thinking and talking about psychological causality in themselves and in others.

A second possibility accounting for better performance on the social inferencing tasks relates to the increasing social experience just mentioned, specifically, the role of situational (or event based) knowledge on false belief tasks. In the logical inferencing tasks, situational knowledge appears less relevant for arriving at a conclusion, whereas linguistic knowledge appears more relevant. Consider the first-order explicit false belief scenario: “This is Pam. She wants to find her cat. Pam’s cat is really in the kitchen. Pam thinks her cat is in the bathroom. Where will Pam look for her cat?” Resolution of the problem depends on attention to a comparison of the key linguistic information that marks the appearance-reality distinction as related to object location in order to derive a conclusion. In contrast, the first- and second-order social inferencing tasks, although they appear to draw on more complex emotional states, are better supported by situational knowledge and depend less on the foregrounding of linguistic knowledge. To illustrate, the explicit false belief task concerned Mrs. Perez wanting to find her son, Juan, because she thought he was lost. The question asked how she would feel when she did not find him on the playground (Juan was really in the park). By drawing on more general situational knowledge of how anyone typically “feels” when not finding someone in the anticipated location, the likelihood is increased of deriving an appropriate pragmatic inference about the mental state that would result.

**Effects of the Prompt Hierarchy**

The patterns of performance showed that the prompt hierarchy appeared to improve the accuracy of pragmatic inferencing for the ASD group, similar to the findings of Rieffe et al. (2000). A combination of the cloze and forced choice procedures enhanced performance only in the social inferencing condition and primarily for the first-order tasks. The fact that prompt use was effective suggested that the participants with ASD were not merely guessing or repeating previous information but, instead, were using the cues offered in some manner. Hence, the use of support indicates that, for the group with ASD, more general situational knowledge of feelings was at least partially encoded in long-term memory and the level of support provided served to facilitate the pragmatic inferences required to resolve a social dilemma in the most plausible way.

Stated another way, prompting was effective when some degree of knowledge existed and the prompts could then draw attention to a specific semantic focus while simultaneously functioning as a verbal support to access that information. This suggests that, for this sample with ASD, certain theory-of-mind capacities existed, albeit in an underdeveloped or even underused state, which then could be accessed with a more favorable response format. On the other hand, the possibility cannot be ruled out that, because individuals with ASD likely process social information in a different way than do typically developing children, they are overly dependent on their personal experience of situations. As a result, they miss the point. Even with the use of the prompts, there is no assurance that their “correct” response holds the same meaning for them as it does for typically developing children. Furthermore, procedures that enhance task performance, like the use of prompts, as Wellman et al. (2001) noted, are not the same as the independent display of what a child knows. The inference that a child’s false belief judgments “reflect robust, deep-seated conceptions of...belief states” (p. 671) is a valid conclusion only when that child needs a minimum of assistance to be successful.

In conclusion, results from this study would indicate that task construction combined with a prompt hierarchy makes a difference relative to disparities that may not be as visible on the standard tasks. For example, the LA group did not perform as well as expected, whereas the group with ASD performed better with support than expected. This outcome was particularly evident for performance on the social inferencing tasks. In other words, when the access of real-world knowledge is supported, the pragmatic inferencing necessary to engage in resolving social paradoxes, such as basic false belief understanding, can be facilitated.

**LIMITATIONS**

At least three limitations to this study can be identified. One reason why the overall performance of the LA group may have been depressed relative to existing developmental data is due to their manner of selection. In contrast to the CA and ASD groups, the LA group was not recruited from school settings, but was differentially recruited through advertising. Thus, the participants who met the inclusion criteria, including the language age criterion, may not have been a representative sample of typically developing

---

Silliman et al.: A New Look 247
children. It is important to note that the matching criteria resulted in a wide range of ages and language age scores; thus, individual variability may have exceeded group homogeneity (Bartsch & Estes, 1996).

The second limitation relates to the appropriateness of standardized measures of lexical recognition, such as the PPVT-III (Dunn & Dunn, 1997), as a matching criterion. These measures are global assessments of vocabulary size only and lack sufficient sensitivity for unraveling relationships between the depth of lexical knowledge and the nature of the conceptual understandings that constitute successful performance on false belief tasks regardless of their mental state content. In this study, neither the multiple meanings of the mental state references that conveyed the physical and psychological states expressed in the two sets of tasks, such as know and feel, nor abilities in false complement processing (e.g., Booth & Hall, 1995; de Villiers & de Villiers, 2000), were assessed in the three groups.

A third limitation might be related to the possibility that the not-own belief tasks (tasks 2 and 6) encourage repetition of the choice offered (i.e., “That’s a good guess. But Judy really feels [the examiner states the opposite of what the participant said previously]. How does Judy feel?”). Although it is a likelihood that repetition may have been a factor for these two tasks, the fact that the group with ASD benefited from prompts primarily in the social inference condition does not support this idea. Future studies should address this issue, however.

In conclusion, two areas merit further clinical investigation. First, an important question concerns how specific educational practices affect the development of social inferencing skills and the appropriate use of language as the propositional medium for the comprehension and expression of beliefs in both the oral and the text domains. Detailed study of classroom interactions could provide insight into how particular curricula and interactional experiences, such as more natural social problem-solving situations, promote or fail to promote theory-of-mind development and its multifaceted expressions in students with ASD. Second, emerging knowledge and skills can be easily overlooked by independent measures of performance. A graduated prompt procedure, more elaborate than the procedure used in this study, can offer successive invitations to try out plausible inferences, while also providing critical information on an individual’s learning potential. In this context, learning potential is defined by how much and what kinds of verbal and nonverbal supports are needed to master a concept and apply it to new situations that also have educational benefits for that individual (Tzuriel, 2001). As the “glimmer” of social inferencing ability arising from this study suggests, going beyond a student’s “inability” in order to access emerging knowledge may provide more realistic profiles of particular differences in performance.

ACKNOWLEDGMENTS

We acknowledge the important contribution of Elizabeth Clark to this study. We express our gratitude, as well, to the students, families, and school programs that made the study possible.

REFERENCES


Bishop, D. V. M. (2000). What’s so special about Asperger syndrome: The need for further exploration of the borderlands of


Received October 22, 2002
Accepted May 15, 2003
DOI:10.1044/0161-1461(2003/020)

Contact author: Elaine R. Silliman, Communication Sciences and Disorders, University of South Florida, 4202 East Fowler Avenue, PCD1017, Tampa, FL 33620-8100. E-mail: silliman@ehuma1.cas.usf.edu
APPENDIX A. LOGICAL INFERENCING SCENARIOS CONSISTING OF THREE FIRST-ORDER TASKS (#1–3), ONE SECOND-ORDER TASK WITH CONTROL QUESTIONS (#4), AND THE PROMPT LEVELS (TASK ORDER REFLECTS THE DEVELOPMENTAL SEQUENCE OF ACQUISITION IN TYPICALLY DEVELOPING CHILDREN)

1. **Inferred Belief** (modified from Wellman & Bartsch, 1988): This is Carlos. This morning Carlos saw his glasses on the shelf, not on the table. Now he wants his glasses. Where will he look for them? (Correct answer: shelf). (If necessary, fill in: Carlos will look on the ______. Forced choice: Will Carlos look on the shelf or will Carlos look on the table?)

2. **Not-Own Belief** (modified from Wellman & Bartsch, 1988): This is Pam. She wants to find her puppy. It might be hiding in the house or outside. Where do you think Pam’s puppy is hiding? That’s a good guess. Pam thinks her puppy is _____. Where will Pam look for her puppy? (If necessary, fill in: Pam will look ______. Forced choice: Will Pam look in the house or will Pam look outside?)

3. **Explicit False Belief** (modified from Wellman & Bartsch, 1988): This is Pam. She wants to find her cat. Pam’s cat is really in the kitchen. Pam thinks her cat is in the bathroom. Where will Pam look for her cat? (If necessary, fill in: Pam will look in the _____. Forced choice: Will Pam look in the kitchen or will Pam look in the bathroom?)

4. **Implicit False Belief** (modified from Sullivan et al., 1994): This is a story about Pam and her Dad. Today is Pam’s birthday and she’s having a big party tonight. Dad is surprising her with a new bike that he has hidden in the living room. See? Here is the surprise bike. Pam and Dad are in the kitchen talking about her birthday. Pam says, “Dad, I really want a new bike for my birthday.” Now remember, Dad wants the bike to be a surprise, so he says, “Sorry, I didn’t get you that. I got you roller blades instead.”

First-Order Question: What does Pam think Dad got her for her birthday? (If necessary, fill in: Pam thinks Dad got her _____. Forced choice: Pam thinks Dad got her a bike for her birthday or Pam thinks Dad got her rollerblades for her birthday.)

Reality Question: What did Dad really get her for her birthday? (If necessary, fill in: Dad really got Pam ___. Forced choice: Dad really got Pam a bike for her birthday or Dad really got Pam rollerblades for her birthday.) Then Pam says to Dad, “O.K., well I’m going over to my friend’s house. I’ll be home later.” On her way out, Pam goes into the living room to get her umbrella because it’s raining. In the living room, she finds her new bike! She thinks to herself, “Yes! Dad did not get me roller blades. He really got me a bike.” Dad does not see Pam go into the living room and find the bike.

Linguistic Contrast Question: Does Dad know that Pam saw her bike in the living room? Later, Pam’s grandmother comes over for the party. Grandma asks Dad, “Does Pam know what you got her for her birthday?”

Ignorance Question: What does Dad say? (If necessary, fill in: Dad says _____. Forced choice: Does Dad say, “Yes, Pam knows what I got her for her birthday” or “No, Pam does not know what I got her for her birthday”?) Now remember, Dad does not know that Pam saw what he got her for her birthday. Then Grandma asks Dad, “What does Pam think you got her for her birthday?”

Second-Order Question: What does Dad say? (If necessary, fill in: Dad says, “Pam thinks I got her _____. Forced choice: Does Dad say, “Pam thinks I got her a bike” or “Pam thinks I got her rollerblades”?)

Justification Question: Why does Dad say that?
APPENDIX B. SOCIAL INFERENCING SCENARIOS CONSISTING OF THREE FIRST-ORDER TASKS (#5–7), ONE SECOND-ORDER TASK WITH CONTROL QUESTIONS (#8), AND THE PROMPT LEVELS

5. **Inferred Belief:** This is Joe. Joe spills ketchup on his shirt. Joe feels upset not surprised. How does Joe’s face look? *(If necessary, fill in: Joe’s face looks ______. Forced choice: Does Joe’s face look surprised or does Joe’s face look upset?)*

6. **Not-Own Belief:** This is Judy. Judy wants to find her puppy. The puppy is hiding from her. Judy might be angry because she can’t find the puppy, or she might be curious about where the puppy is hiding. How do you think Judy feels? *(Child answers). That’s a good guess but Judy really feels ___(Different feeling than child chose). How does Judy feel? (If necessary, fill in: Judy feels ___. Forced choice: Does Judy feel angry or does Judy feel curious?)*

7. **Explicit False Belief:** This is Mrs. Perez. She wants to find her son, Juan. Mrs. Perez is worried because Juan is lost. She will be relieved when she finds her son. Juan is really at the park, but Mrs. Perez thinks he is at the playground. How will Mrs. Perez feel when she does not find Juan at the playground? *(If necessary, fill in: Mrs. Perez will feel ___. Forced choice: Will Mrs. Perez feel relieved when she does not find Juan at the playground or will Mrs. Perez feel worried when she does not find Juan at the playground?)*

8. **Implicit False Belief:** This is a story about Frank and his Dad. Today is Frank’s birthday. Frank wants to go to a baseball game for his birthday. He does not want a surprise party for his birthday. Frank hates surprise parties. He gets embarrassed when everyone looks at him and yells “Surprise!” Frank’s Dad is giving him a surprise birthday party. Dad bought Frank balloons and a birthday cake and hid them in the living room. Dad does not know that Frank gets embarrassed at surprise parties. Dad thinks Frank would be glad to have a surprise party. Frank and Dad are in the kitchen talking about his birthday. Frank says, “Dad I really want to go to a baseball game for my birthday.” Now remember, Dad wants the party to be a surprise, so he says, “Frank, that’s a good idea. I will think about it.”

---

First-Order Question: What does Frank hope that Dad will do for his birthday? *(If necessary, fill in: Frank hopes that Dad will _____. Forced choice: Does Frank hope that Dad will take him to a baseball game for his birthday or does Frank hope that Dad will give him a surprise party for his birthday?)*

Reality Question: What did Dad really do for Frank’s birthday? *(If incorrect, say: “But remember, Dad plans to surprise Frank with the party.” (If necessary, fill in: Dad really _____. Forced choice: Did Dad really give Frank tickets for a baseball game for Frank’s birthday or did Dad really plan a surprise party for Frank’s birthday). Then Frank says to Dad, “Great, I’m going to tell my friend next door, I’ll be home later.” On his way out, Frank sees the balloons and cake hidden in the living room. He thinks, “Oh no! Dad is giving me a surprise party; we are not going to the baseball game.” Remember, Dad does not see Frank go into the living room and find the balloons and cake.)*

Linguistic Contrast Question: Does Dad know that Frank is disappointed? *(If necessary, fill in: n/a. Forced choice: Yes, Dad knows Frank is disappointed or No, Dad does not know Frank is disappointed. If wrong, say: “But remember, Dad does not see Frank go into the living room and find the balloons and cake.”)* Later, Frank’s grandmother comes over for the party. Grandma asks Dad, “Does Frank like surprise parties?”

Ignorance Question: What does Dad say? *(If necessary, fill in: Dad says ___. Forced choice: Does Dad say, “Yes, Frank likes surprise parties” or “No, Frank does not like surprise parties”?) Now remember, Dad does not know that Frank gets embarrassed at surprise parties. Then Grandma asks Dad, “How will Frank feel about having a surprise party?”

Second-Order Question: What does Dad say? *(If necessary, fill in: Dad will be ____. Forced choice: Does Dad say, “Frank will be glad to have a surprise party” or “Frank will be embarrassed to have a surprise party”?)*

Justification Question: Why does Dad say that?