Developmental Psychology
1998, Vol. 34, No. 5, 1026-1037

0012-1649/98/$3.00

Understanding Mind and Emotion: Longitudinal Associations With Mental-State Talk Between Young Friends

Claire Hughes and Judy Dunn
Institute of Psychiatry

Developmental changes in children's understanding of mind and emotion and their mental-state talk in conversations with friends were examined in a longitudinal study of 50 children (*M* age at each time point = 3 years 11 months, 4 years 6 months, 5 years 0 months). Significant and related improvements over time were found for both theory-of-mind task performance and affective perspective taking. Associated with these cognitive developments were quantitative and qualitative changes in children's references to mental states in their conversations with friends. Individual differences in theory of mind, emotion understanding, and mental-state talk were strikingly stable over the 13-month period. Although there were no gender differences in children's task performances, girls showed more frequent and more developed mental-state talk than boys.

Developmental changes in young children's understanding of mental states have been researched intensively in the past 15 years. Despite the number of studies in the field, very few have had a longitudinal developmental design. As a result, little is known about the stability of individual differences in children's mentalizing skills. Longitudinal studies are also needed to explore relations between developmental trajectories in different domains. For example, understanding belief and understanding emotion are linked conceptually but may show contrasting antecedents and sequelae (J. Dunn, 1995). Similarly, researchers have recently begun to explore gender differences in children's early social understanding; however, it is not clear whether these differences can be explained by gender differences in language development (Happé, 1995) or in peer relationships (Ruble & Martin, 1998). In this article, these three topics are addressed through a 1-year longitudinal study of 25 pairs of friends.

Stability of Individual Differences in Understanding Belief and Its Correlates

The ever-growing number of experimental studies of children's understanding of mind reflects a common consensus that understanding mental states is essential for everyday social relations. Yet surprisingly little is known about the real-life correlates of individual differences in children's performance on tests of mental-state understanding. However, four recent observational studies have shown a significant correlation between false-belief performance and both the quality (Astington & Jenkins, 1995; Taylor & Carlson, 1997) and quantity (Hughes & Dunn, 1997; Youngblade & Dunn, 1995) of children's pretend play with friends. Other observational studies have shown a correlation between false-belief task performance and both the connectedness of children's communication (Slomkowski & Dunn, 1996) and the frequency and sophistication of their talk about mental states with friends (Brown, Donelan-McCall, & Dunn, 1996; Hughes & Dunn, 1997).

Investigations such as those above not only suggest that individual differences in mental-state awareness are meaningful for children's social lives, but also demonstrate the utility of combining experimental and observational methods. At the same time, the issue of external validity appears especially salient in observational studies, because subtle changes of context have been associated with striking differences in talk and behavior (J. Dunn, Brown, & Beardsall, 1991; Gottman, 1986). These contrasts suggest that laboratory-based research in which children are characterized as theory-of-mind "passers" or "failers" may be misleading. This issue is highlighted by a recent study indicating poor test–retest reliability for false-belief tasks (Mayes, Klin, Tercyak, Cicchetti, & Cohen, 1996). Examining the stability of individual differences in false-belief task performance (and its correlates) was therefore an important aim of the present study. Highly stable differences between children were predicted, primarily on the basis of findings from an early study by J. Dunn, Brown, Slomkowski, Tesla, and Youngblade (1991) in which family talk about feelings and cooperative interactions between siblings at 33 months were found to be related to children's social understanding 7 months later.

Relations Between Understanding Belief and Emotion and Children's Mental-State Talk

Studies of children's understanding of emotions indicate that individual differences are associated with differences in pro-

---

Claire Hughes, Medical Research Council Child and Adolescent Psychiatry Unit, Institute of Psychiatry, London, England; Judy Dunn, Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, London, England.

This work was conducted with the support of National Institutes of Health Grant HD 23158. We thank all children and staff at the following South London and Kent nurseries and schools for their participation in the study: Crawford, Dulwich Village, Dulwich Wood, Grove, Heber, Langbourne, Mayfield, St. Alban's, St. Bartholomew's, St. Francis's, St. George's, St. Luke's, Sumner Nursery, and Sunnyhill. We also thank Francesca Happé and Tom O'Connor for their helpful comments on a previous version of this article.

Correspondence concerning this article should be addressed to Claire Hughes, Medical Research Council Child and Adolescent Psychiatry Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, England. Electronic mail may be sent to clare.hughes@iop.bpmf.ac.uk.

social behavior (Denham, McKinley, Conchoud, & Holt, 1990) and show some stability over time (Brown & Dunn, 1996). However, these patterns may not generalize to children's understanding of mental states such as belief. For example, using a Pennsylvania sample of 40-month-old children, J. Dunn (1995) found that individual differences in performance on tests of emotional perspective taking and false-belief explanation were unrelated. Indeed, at follow-up when the children were 6 years old, early understanding of emotions was related to positive perception of peer experiences, whereas early false-belief understanding was related to negative initial perceptions of school, greater sensitivity to teacher criticism, and harsher self-judgments of performance (J. Dunn, 1995). A second aim of this study was to track children's developing understanding of emotion and belief to explore Dunn's hypothesis that developments in each domain have differential sequelae.

A third aim of the study was to investigate relations between children's interactions with friends and their developing understanding of mind. An alternative and tempting interpretation of J. Dunn, Brown, Slomkowski, et al.'s (1991) findings of an association between cooperative sibling interaction at 33 months and false-belief performance at 40 months is that peer interactions have a facilitatory effect on sociocognitive development (see Hartup, 1996). That is, sibling interactions could provide an enriched context for children's learning about others' minds. Support for the association between sibling interaction and children's developing understanding of mind can be found in three recent independent studies. Perner, Ruffman, and Leekam (1994) reported that children from large families showed an accelerated rate of success on standard false-belief tasks. Jenkins and Astington (1996) replicated this effect but found that the beneficial effect of siblings was only significant for verbally less-able children. Finally, Lewis, Freeman, Kyriakidou, Maridaki-Kassotaki, and Berridge (1996) found facilitatory effects of close contact with not only siblings but also members of the extended family. The conclusion from these studies is that even if mentalizing ability is innately specified (Leslie, 1987), children's social experiences may well play an important role in its development.

Another lesson to be learned from longitudinal research is that there may be important developmental changes in when and why children talk about mental states. For example, at 33 months of age, young children's talk about inner states is predominantly with their mothers (and often in the context of disputes with siblings), whereas at 47 months of age, children show more frequent mental-state talk with their siblings or friends, and this talk is often in the context of cooperative play (Brown et al., 1996). In itself, this developmental shift is extremely interesting; the findings are also useful at a practical level, as they enable researchers to focus on specific contexts that are likely to be rich "seams" for investigating children's conversations about mental states (e.g., Hughes & Dunn, 1997).

A focus on friends also provides a bridge between cognitive perspectives on children's development and the growing body of social research into the developmental significance of children's friendships. With the exception of Howes's (1988) studies of preschoolers, much of this work has been done with school-age children and adolescents. Yet forming friendships is a major social achievement for preschool-age children, and one that may well be tied to their growing understanding of others' beliefs, intentions, and desires. In addition, numerous studies have shown that conversations between friends are more vigorous, mutually oriented, and elaborated than those between nonfriends (Newcomb & Bagwell, 1995). For these two reasons, conversations between friends offer a privileged view on how young children apply their developing sociocognitive skills to real-life interactions with close others.

Gender Differences in Social Understanding and Mental-State Talk

A final topic of interest in the present study stems from recent findings that suggest gender differences in cognitive and affective perspective taking (Brody, 1985; Happé, 1995). One possibility is that the differences between boys' and girls' performances on tests of understanding mind and emotion reflect gender differences in language development (e.g., Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). Alternatively, these differences in social understanding may relate to differences in the quality of boys' and girls' early social relationships with peers: Friendships between girls may be more intimate and mutually oriented, and so lead to more rapid development in their understanding of others.

Method

Participants

At the start of the study, 50 children (25 pairs of friends) from four local inner-city nursery schools were recruited: 25 boys and 25 girls; 25 Caucasian children and 25 children of African or Caribbean origin. All children spoke English as their first language at home. Participants included 10 boy–boy pairs, 10 girl–girl pairs, and 5 boy–girl pairs, and ranged in age from 3 years 3 months to 4 years 7 months (M = 3 years 11 months, SD = 5 months). Pairs of friends were identified by class teachers and were included if also nominated by a second member of staff. The mean duration of the children's friendship prior to testing was 9 months (range = 2 months to 3 years). Friends met every day at their nursery, and on average met 2–4 times a month outside the nursery.

As the children were followed up over the year, most of the group moved from nursery to primary schools. Although many pairs of friends were separated by this transition to school, several friendship dyads continued to play together outside school, either because they lived very near each other or because their mothers were friends. As a result, it was possible to film 19 (75%) of the original 25 friendship dyads at all three time points over the course of the year, even if 1 or 2 of these film sessions took place in one friend's home rather than in the nursery. The remaining 6 pairs were filmed with new friends in their new schools. Although the new children recruited to the study over the course of the year were included in all the testing sessions, for the sake of simplicity the data reported here are restricted to results from the original group of 50 children. All of these children were seen at all three time points, so there was no dropout from the study.

Each of the four nurseries used for initial recruitment served a predominantly low-income catchment area. Parental occupation of participants in this study reflected this distribution, although families ranged from those in which both parents were unemployed (8 cases) to those in which both parents were professionally employed (7 cases). Fathers' occupations for the previous 2-years were summarized as follows: unemployed (13); domestic worker (7); manual worker (7); cab driver (5); clerical worker (5); accountant (8); teacher (1); engineer (3); and doctor (1).

Experimental Design

The aim of the study was to observe friendship pairs and administer a set of tasks tapping children's understanding of mind and emotion at three time points evenly spread over a 1-year period. School holidays disrupted the scheduling somewhat, and so the data were in fact collected at intervals of 7 and 13 months from the start of the study. The mean age of the group was 3 years 11 months at Time 1; 4 years 6 months at Time 2; and 5 years exactly at Time 3 (SD = 4.7, 5.2, and 4.9 months, respectively).

At each time point, friends were filmed for 20 min in dyadic play in a quiet room of the nursery or school, equipped (by the experimenter) with a big box of toys and dressing-up materials (see *Materials* section), and a cine-camera mounted on a tripod in one corner. One week later, children were tested individually for their understanding of mind and emotion. Understanding of mind was tested using three parallel task sets that were counterbalanced across time points and participants. At each time point, children were given two standard false-belief prediction tasks, one emotion false-belief prediction task, three false-belief explanation tasks, and two deception tasks. Understanding of emotion was tested using Denham's (1986) affective labeling and affective perspective-taking tasks at Times 1 and 2, and Gordis, Rosen, and Grand's (1989) mixed emotion task at Time 3. In addition, at Times 2 and 3 children were interviewed about their real-life understanding of causes of emotion for themselves and close others. Their responses at each time point were extremely similar (see J. Dunn & Hughes, 1998), and so the data presented here are from Time 2 only. Background measures of children's verbal and nonverbal abilities were obtained at Time 1 using the British Picture Vocabulary Scale (L. M. Dunn, Dunn, Whetton, & Pintilie, 1982) and all six performance subtests of the McCarthy Intelligence Scales (McCarthy, 1972).

Materials

Dressing-Up Materials and Role-Play Toys

Items included one fireman's hat; one pirate hat, jacket, toy hook, and cutlass; one cowboy hat; one scary mask; one fairy mask; one crown; one ballet tutu; one toy handbag with money, sunglasses, and jewels; two ragdolls and two Sesame Street handpuppets; one toy police set with two hats, two Walkie-Talkie phones, two handcuffs and two badges; one toy doctor's case with stethoscope, thermometer, otoscope, tendon-hammer, bandage, and syringe; one toy cookset with cooker, blender, pans, plates, cups, cutlery, and food; one toy toolbox with multi-colored torch, spirit level, tape-measure, pliers, saw, and hammer. (At the second and third time points, these dressing-up materials were renewed, although the themes for the toys were unchanged.)

Task Materials

Items included two clown dolls (one red, one blue); two visually distinct pots and a small ball for the unexpected location false-belief prediction task; one prototypical candy tube (Smarties/M&Ms) containing a pencil for the deceptive contents false-belief prediction task; three prototypical boxes (BandAid box, egg box, cereal box); three plain boxes and six handpuppets (three boy puppets, three girl puppets) for the two unexpected location false-belief explanation tasks and one true-belief filler task; and one "lock-a-block" box (200-mm cube) with large plastic child-friendly key, one princess handpuppet, one pirate handpuppet, and chocolate gold coins for the sabotage/deception task.

Procedure

Theory-of-Mind Tasks

At each time point, children were presented with eight false-belief tasks (five prediction tasks, three explanation tasks) and two deception tasks. These tasks are summarized briefly below.

False-belief prediction tasks. These tasks provide a crucial test for children's understanding of the representational and fallible nature of belief. At each time point, children were presented with two unexpected location tasks, two unexpected identity tasks, and one emotion false-belief task. In the unexpected location task (adapted from Wimmer & Perner, 1983), the children were told a story involving a target object (e.g., a ball), two opaque containers (e.g., a box and a basket), and two puppet characters. In this story, one puppet moved the target object while the other puppet was absent. The uninformed puppet then returned and expressed a desire for the target object. Children were asked to predict in which of the two containers the puppet would look. Children were also asked a reality control question (e.g., "Where is the ball really?") and a memory control question (e.g., "Where did the puppet leave the ball?"), and they were only coded as successful if they gave correct responses to all of the control questions and to the test question on each task.

In the unexpected identity tasks, children were either shown a box with deceptive contents (e.g., a biscuit box that contained a toy animal) or told a story from a peep-through pop-up book, in which the final page included an element with a deceptive identity (e.g., a spot on a snake's back, which through the hole on the previous page had looked like an eye). For the peep-through storybook tasks, children were asked what they could see both before and after turning the last page (belief-control and reality-control questions). Turning back to the penultimate page, children were introduced to a naive puppet and were asked to predict what the puppet would think the picture was before looking at the last page. For the deceptive contents task, both before and after opening the box, children were asked what they thought was inside. The box was then closed over, and the children were again asked to predict what a naive puppet would think was inside the box. Children were only coded as successful if they gave correct responses to all of the control questions and to the test question on each task.

At each time point, children also received one of three emotion false-belief prediction tasks, based on the story developed by Harris, Johnson, Hutton, Andrews, and Cooke (1989). This task is designed to test children's understanding of a character's false belief and also how the character will feel as a result of that false belief. The three stories used in the present study involved a nice surprise, a nasty surprise, or a scary surprise, and these were counterbalanced across time points. In the nasty surprise story, a character (naughty Micky the monkey) exchanged his friend's favorite drink (Coke) for a disliked drink (milk) while his friend was away. In the nice surprise story, a character exchanged a disliked food for a liked food, again in the friend's absence. In the scary surprise story, each character attempted to scare the other by hiding behind a tree and pretending to roar like a lion (but while one character hid himself very well, the other could be seen from his hiding place). Children were credited with success on each story if they correctly stated what the friend thought he was getting to eat/drink (or who each friend thought was behind the tree for the scary story), and what he was really getting to eat/drink (or who was really behind the tree). Children were also given an extra point if they could also predict how the friend would feel in each story (possible score = 0–2).

False-belief explanation tasks. These tasks were on the puppet task developed by Bartsch and Wellman (1989) and involved prototypical boxes and unexpected location stories. For each story, children were presented with a prototypical box (e.g., an egg box) and a matching plain box and were shown that the prototypical box was empty, whereas the plain box contained the target object (e.g., a toy egg). A filler story, in which the prototypical box contained the target object and the plain box was empty, was also included to prevent children guessing that the plain box (rather than the prototypical box) would always be full. For each story, the children were introduced to a puppet "who has not seen these boxes before," who expressed a desire for the target object and walked toward the prototypical box. Children were asked, "Why is (puppet) looking there?" If no belief-based answer was given, children were prompted with the question, "What does (puppet) think?" Children were rated as successful on each story if they made either a spontaneous or a prompted reference to a mistaken belief in explaining the story character's action (possible score = 0–3).

Deception tasks. In addition to the above story vignettes, two deception tasks were used at each time point. The first of these was Sodian and Frith's (1992) one-box puppet deception game. This four-condition task involves not only cooperative and competitive trials but also physical versus verbal conditions (all four conditions were counterbalanced across children). In the physical condition, children are shown a box that can be locked using a large plastic key and are introduced to a nice princess puppet and a mean burglar puppet. A chocolate gold coin is hidden in the box on each trial, and the children are instructed to help the princess (cooperative trial) and to make it difficult for the burglar to find the gold coin (competitive trial). In the verbal condition, the key is removed, and the princess and burglar puppets are made to appear from a distance and say, "I wonder if the box is locked or open? If the box is locked, I won't bother making the journey, it's such a long way." Again, the children were encouraged to help the princess but to make it difficult for the burglar, and they were asked what they would tell the puppet: "Are you going to say the box is locked or open?" (word order was counterbalanced across children). Children scored 1 point for passing both trials in both the verbal and the physical condition, so possible scores ranged from 0 to 2 points.

The second task was a penny-hiding game, familiar to most preschoolers and used recently in studies of children with autism (Baron-Cohen, 1992; Oswald & Ollendick, 1989). The experimenter hid a coin behind her back and, bringing both hands forward with the coin concealed in one hand, asked the child to guess which hand held the coin. This was repeated for three trials, after which the experimenter announced that it was now the child's turn to hide the coin. On each of the three test trials with the child as hider, children scored 1 point if they fulfilled three criteria: (a) invisible displacement of the coin, (b) both hands presented for guessing, and (c) coin concealed throughout. Possible scores on this task therefore ranged from 0 to 3 points.

Theory-of-mind aggregate. From the five types of theory-of-mind task (standard false-belief prediction, emotion false-belief, false-belief explanation, sabotage/deception, and penny hiding), an aggregate theory-of-mind score was computed by scaling each score to a 0–6 range (to balance the contribution from each type of task). The maximum score was therefore 30 points at each time point. To avoid loss of data, we rated children who failed control questions on any particular task as having failed that task. Cronbach's alpha for the theory-of-mind aggregate was .68 at Time 1, .75 at Time 2, and .59 at Time 3, indicating fair to good scale reliability.

Emotion Understanding Tasks

At Times 1 and 2, children's affective labeling and affective perspective-taking abilities were assessed using the procedures developed by Denham (1986). (Children's scores on these tasks approached ceiling at Time 2, and so the tasks were not repeated at Time 3.)

Affective labeling. Children were shown four felt faces portraying happy, sad, angry, and frightened expressions. They were asked to identify the emotion portrayed in each face, first expressively, by naming, and then receptively, by pointing to the expression named by the experimenter. The code for scoring on this task was exactly the same as that used by Denham (1986). Children received 2 points for correct naming or pointing; 1 point for identifying the correct valence but mistaking the specific emotion (e.g., calling the frightened face sad); and 0 points for giving a wrong valence emotion (e.g., calling the frightened face happy) or for failing to provide a response. The maximum possible score for affective labeling was therefore 16 points (expressive and receptive identification of four emotions).

Affective perspective taking. In this task, puppets (with blank faces) were used to enact 17 vignettes portraying situations in which the protagonist felt happiness, sadness, anger, or fear, such as going to the zoo, seeing a parent off on a trip, having a toy hidden by a sibling, or having a bad dream, respectively. Each of the stories was acted out with vocal and facial cues for the puppet's feelings. Eight of the vignettes were unambiguous, in that the puppet's emotion was what most people would be expected to feel in that situation. The other 9 vignettes were ambiguous, in that the puppet's emotion differed from the emotion previously predicted by the mother for the child. At the end of each story, the children were asked how the protagonist felt. Children could respond verbally or nonverbally by selecting the appropriate felt face to stick onto the puppet (children were credited with their best answers in the rare case of giving contradictory verbal and nonverbal responses). Responses were scored as in the affective labeling task (maximum possible total score = 34 points). Following Denham (1986), an aggregate score for affective labeling and affective perspective taking was created by combining the children's scores on the two sets of tasks (maximum possible score = 50). Cronbach's alpha for the aggregate score was .85 at Time 1 and .68 at Time 2, indicating good scale reliability.

Emotion interview. At Time 2, children were also interviewed about the real-life causes of happiness, sadness, anger, and fear for themselves, their friends, and their mothers. This interview was based on that used by Cassidy, Ross, Butkovsky, and Braungart (1992) and is reported more fully in J. Dunn and Hughes (1998). In brief, children were shown four picture cards representing a happy face, a sad face, an angry face, and a scared face. After the child had correctly identified what emotion each face represented, the first face was shown and the child was asked four questions: (a) What kinds of things make you feel this way? (b) Can you give me an example of a time you felt this way . . . then what happened? (c) Let's pretend you saw (friend's name) looking this way—why do you think she/he might be looking like that? (d) Let's pretend you saw your Mum looking this way? Why do you think she might be looking like that? The procedure was repeated for each of the four emotions. The children's responses were tape-recorded and transcribed verbatim following the session. They were then coded in terms of *theme, agent* (referred to as cause), and *adequacy.* The results presented in this article are from the adequacy coding. Responses were coded on a 0–4-point scale: 0 = *no response, refusal, don't know*; 1 = *poor response* (including irrelevant remarks, failure to understand causal nature of question, or cause suggested that is very unlikely to provoke emotion in question); 2 = *adequate response* (including one word or simple clause only, but appropriate and plausible response; e.g., "monsters" as a reply to the question of what made a friend scared); 3 = *good response* (including relevant appropriate sentence response or two or more adequate responses); 4 = *excellent response* (including elaborate response or evidence of insight; e.g., mention of mixed emotions, transience of emotional experience). Examples for happiness in self follow: "Fireworks. They're going to be at the 1 o'clock club. I'll be very happy because I haven't seen it before." or "When your Mum brings you a present as it's your birthday, and then you feel sad again because the birthday is finished." Agreement between the two coders was assessed for 25 children. The value of Cohen's kappa for adequacy ratings was .87, indicating excellent interrater reliability. Cronbach's alpha for the overall adequacy score was .91, indicating excellent scale reliability.

Mixed-emotions task. At Time 3, children's understanding of mixed emotions was assessed using the procedures developed by Gordis et al. (1989). The first part of this task consists of three short vignettes, each involving a protagonist who has mixed feelings about a situation (e.g., feeling happy about winning a race against a friend, but also feeling sad because that friend fell down in the race). Children were asked to give a reason for each emotion expressed by the story character. In the next three vignettes in the second part of the task, the mixed emotions of the story characters are not made explicit. Here, children were asked to identify how the protagonist (X) feels and to explain that emotion. In this part of the task, children were assisted by the prompt "Does X feel anything else? Why does she/he feel that way?" For both parts of the task, children were credited with 0–2 points for each vignette. To receive

2 points, children had to provide an adequate explanation for why a character might simultaneously experience positive and negative feelings about a given situation. Children who provided adequate explanations for only one emotion or explained successive positive and negative feelings were credited with 1 point; all other answers scored zero. In the third part of the task, children were asked to give an account of a time in which they had themselves experienced mixed emotions. However, few children in the study were able to do this, and so aggregate scores were calculated from the first two parts only (maximum score = 12). A Cronbach's alpha of .53 was obtained for this aggregate score, indicating fair scale reliability.

Coding Mental-State Talk

In transcribing the videotapes of the children's dyadic play sessions, we defined a *conversational turn* as all of one child's utterances bounded by the utterances of the friend. The total number of child utterances at Times 1, 2, and 3 was 4,748, 6,740, and 8,220, respectively (grand total = 19,746 utterances). The analyses focus on conversational turns that included a term denoting a mental state (n = 167, 332, and 421, at Times 1, 2, and 3, respectively). Reliability of transcription was checked by Claire Hughes, who transcribed a 5-min section of each video (i.e., 25% of all transcripts). Good agreement was found for number of speaker turns, $r(49) = .96$, and number of mental-state turns, $r(49) = .95$.

Coding of mental-state terms was based on earlier studies of children's discourse (Shatz, Wellman, & Silber, 1983; Furrow, Moore, Davidge, & Chiasson, 1992; Brown et al., 1996). The terms included all referred to cognitive mental states (e.g., know, think, pretend, forget, remember, wonder, imagine, dream) and did not include references to desires or feelings, because these states are not explicitly representational. The functional meaning of each mental-state term was coded following Shatz et al.'s system, using the modifications described in Brown et al. (1996) to include "conversational" as well as "genuine" uses of mental-state terms. The four categories derived from the transcripts are as follows:

1. *Genuine mental reference* included all turns in which the speaker referred to his or her own or another's thoughts, beliefs, memories, and so on (e.g., "Do you think Captain Hook could be a policeman?") The phrase "I don't know" without a predicate complement was also included in this category. Contrastives, in which children explicitly contrasted a belief with reality or with another person's belief, were also included in this category; however, these contrastives were extremely rare (only one or two examples at each time point).
2. *Modulation of assertion* included mental-state terms used to strengthen or weaken an assertion (e.g., "It's Casper the ghost, I think") and acknowledgments of the other child's utterance (e.g., "Yes, I know").
3. *Directing interaction* included turns in which the child introduced an activity with a mental-state term (e.g., "Let's pretend we're pirates").
4. *Other* included clarifications (e.g., "Do you mean this one?), nouns, adverbs, and adjectives (e.g., "It's not a real shark, it's only a pretend one).

The *referent* of a mental-state term was coded into three categories: self, other (usually the child's friend), and child plus friend (e.g., "We think it's a dragon"). These last two categories were later collapsed together, to determine total frequency of mental terms that included the other. The *pragmatic context* of the utterance was also categorized in three ways: (a) self-interest (e.g., "I think Mummies wear high heels, so I'll have these"); (b) neutral commentary (e.g., "I think the blue one is best"); and (c) shared interest (e.g., "We're pretending we're cooking, aren't we?"). Reliability was checked by double-coding 20% of the transcripts at each time point. Excellent interrater reliability values were obtained (all Cohen kappa values > .80).

Results

Background Measures of Participants' Ability

At Time 1, the group mean standard index of verbal ability on the British Picture Vocabulary Scale was 120.92 (SD = 15.90). This index is normed around a mean of 100 (SD = 15). The high mean verbal ability for the group reflects the fact that children were recruited on the basis of their close friendships: It is likely that young children who develop such early friendships are verbally more able than their peers, a point that is worth recalling in considering the main results of this study. In contrast, the group mean standard index of nonverbal ability was 54.2 (SD = 10.5), close to the population norm 50 (SD = 10), suggesting age-appropriate nonverbal abilty for the group.

Developmental Change and Stability of Individual Differences

Developmental Change in Task Performances

For understanding emotion, the group mean aggregate Denham (1986) score was 36.90 (SD = 8.44) at Time 1 and 43.29 (SD = 5.10) at Time 2. Children's performances at Time 2 approached ceiling (max. score = 50), and so the Denham task was not included at Time 3. For understanding belief, the numbers of successful children on each type of task (false-belief prediction/explanation and deception) at each time point are summarized in Table 1, together with group mean aggregate scores. Significant increases over time in the number of successful participants were found for both false-belief prediction and deception tasks but not false-belief explanation tasks (see Table 1 for chi-square values). However, half of the group were already successful at false-belief explanation at Time 1, suggesting an earlier time course of improvement for this kind of task.

Within-child developmental changes in aggregate scores for understanding emotion and belief were examined in separate repeated measures analyses of variance (ANOVAs) with age, verbal ability, and nonverbal ability at Time 1 treated as covariates. Children's aggregate scores for understanding emotion increased significantly over time, $F(1, 98) = 51.7$, $p < .001$, and both age and nonverbal ability showed significant covariance with Time 1 Denham scores (β = .36 and .38, respectively, $p < .01$ for both). Within-child improvements in theory-of-mind scores were also highly significant, $F(2, 147) = 19.4$, $p < .001$, and again both age and nonverbal ability showed significant covariance with theory of mind (β = .47 and .41, respectively, $p < .001$ for both). No gender effects were found for aggregate scores of emotion or belief understanding, and there were no significant interaction terms. The lack of covariance between verbal ability and understanding belief and emotion is puzzling but probably reflects the generally high verbal ability of this group.

Developmental Change in Mental-State Talk

Almost all of the children (49 out of 50) showed a steady logarithmic increase in their rate of mental-state talk across the three time points. The nature of mental-state reference (functional category, referent, and pragmatic context) is summarized

Table 1
Number of Successful Participants for Each Theory-of-Mind Task and Aggregate Scores at Each Time Point

Task	Criterion for success	Time point 1	Time point 2	Time point 3	$\chi^2(1, 50)$	$F(2, 49)$
Predict action from false belief	Pass both tasks	18	23	33	9.3**	
Predict emotion from false belief	Pass both emotion and false-belief question	10	19	31	18.5**	
Explain action from false belief	Pass ⅔ tasks	26	32	31	1.7	
Penny hiding	Pass ⅔ trials	25	34	43	14.9**	
Deception	Pass all 4 conditions	28	43	—[a]	16.4**	
Theory of mind aggregate score						19.4**
M		14.3	18.7	21.4		
SD		7.3	8.1	6.6		

[a] A ceiling effect was obtained at Time 2, so this task was not given at Time 3.
** $p < .01$.

for the group at each time point in Table 2. To adjust for skewness, we conducted repeated measures ANOVAs[1] on log-transformed data. The first ANOVA examined within-child effects of time (Times 1, 2, and 3) and functional category of mental-state terms (genuine reference, modulation of assertion, and directing activity), as well as between-child effects of gender. The main effects of time and function were both significant, $F(2, 147) = 9.6$ and 41.4, respectively, $p < .001$ for both, and there was a significant Time × Function interaction, $F(4, 145) = 17.6$, $p < .001$, as well as a marginally significant Gender × Function interaction, $F(3, 146) = 2.9$, $p = .06$. Post hoc analyses showed that genuine mental-state reference was more frequent than conversational or colloquial use of mental-state terms at all time points; this difference increased over time and was marginally more pronounced for girls than for boys.

Table 2
Mean Frequencies of Mental-State (MS) Talk at Each Time Point

Variable	Time point 1	Time point 2	Time point 3	$F(2, 147)$
Speaker turns/hr				16.7**
M	252.0	254.8	351.3	
SD	78.8	131.2	114.3	
MS terms/hr				13.5**
M	10.0	18.4	22.1	
SD	10.1	21.5	19.9	
MS terms/turn				8.8**
M	0.04	0.07	0.06	
SD	0.04	0.08	0.05	
Genuine MS reference/hr				22.3**
M	3.5	12.8	14.3	
SD	5.0	17.9	15.3	
Modulatory MS reference/hr				5.9*
M	1.9	1.8	3.9	
SD	3.0	2.9	4.5	
Directing MS reference/hr				3.2*
M	3.2	2.4	1.1	
SD	5.4	1.8	2.2	
Reference to own MS/hr				28.4**
M	7.1	8.1	11.0	
SD	8.3	8.9	9.9	
Reference to other/shared MS/hr				15.9**
M	3.0	10.1	10.7	
SD	4.7	15.3	14.4	
Self-interested MS reference/hr				2.4
M	1.6	3.9	1.9	
SD	3.4	8.9	5.4	
Neutral MS reference/hr				0.4
M	5.9	5.8	3.7	
SD	6.0	8.1	5.3	
Shared-interest MS reference/hr				41.0**
M	2.4	8.0	17.0	
SD	4.1	10.8	7.1	

* $p < .05$. ** $p < .01$.

A second ANOVA was conducted to examine within-child effects of time (Times 1, 2, 3) and referent for the mental-state term (own vs. other/shared), as well as between-child effects of gender. The results were similar to those above: Significant main effects were found for both time, $F(2, 147) = 13.7$, $p < .001$, and referent, $F(1, 98) = 6.7$, $p < .01$, with a strong Time × Referent interaction, $F(3, 146) = 39.2$, $p < .001$. Post hoc analyses showed that children referred to their own mental states significantly more often than to others' or shared mental states at Times 1 and 2, but this difference disappeared at Time 3.

A third ANOVA was conducted to examine within-child effects of time (Times 1, 2, 3) and pragmatic context of mental-state reference (self-interest, neutral commentary, and shared interest), as well as effects of gender. Again both time and context showed significant main effects, $F(2, 147) = 9.1$ and 35.3, respectively, $p < .001$ for both, and there was a strong Time × Context interaction, $F(4, 145) = 26.8$, $p < .001$. Post hoc analyses showed that at Time 1, most mental-state reference was made in the context of neutral commentary, but at Time 2, and especially Time 3, most mental-state terms occurred in the context of shared interest.

Stability of Individual Differences

Table 3 shows the full correlations (at and between each time point) for both aggregate and individual task scores for belief and emotion understanding, as well as for hourly rates of mental-state talk. When aggregate theory-of-mind scores were compared over time, individual differences were very stable, $r(49)$

[1] These analyses were also run with effects of initial age and ability covaried out, with no change in results or significance levels.

Table 3
Pearson Correlations Between Task Performance and Mental-State Talk at Each Time Point

Variable	1	2	3	4	5	6	7	8	9
1. T1 ToM total	—								
2. T1 1st-order FB	.70**	—							
3. T1 emotion FB	.69**	.47**	—						
4. T1 explain FB	.50**	.44**	.30*	—					
5. T1 penny	.45**	.11	.16	−.13	—				
6. T1 deception	.47**	.15	.16	−.02	.30*	—			
7. T1 Denham	.39**	.41**	.16	.60**	−.09	−.03	—		
8. T1 MS/hr	.21	.26	−.02	.29*	.22	−.10	.33*	—	
9. T2 ToM total	.51**	.45**	.25	.54**	.13	.07	.64**	.30*	—
10. T2 1st-order FB	.28*	.28*	.09	.46**	.06	−.01	.64**	.33*	.79**
11. T2 emotion FB	.40**	.46**	.30*	.46**	.06	.02	.46**	.19	.71**
12. T2 explain FB	.45**	.42**	.24	.45**	.08	−.04	.49**	.28*	.81**
13. T2 penny	.19	.08	.13	.32*	−.03	−.14	.37*	.09	.63**
14. T2 deception	.16	.47**	.00	.09	−.01	.29*	.27	.25	.41**
15. T2 Denham	.47**	.55**	.29	.45**	.00	−.02	.71**	.36**	.51**
16. T2 Cassidy	.23	.26	.09	.46	−.10	−.12	.40**	.16	.52**
17. T2 ms/hr	.41**	.45**	.30*	.53**	−.09	.09	.55**	.44**	.52**
18. T3 ToM total	.34*	.20	.11	.38**	.09	.22	.54**	.45**	.58**
19. T3 1st-order FB	.12	.01	.09	.36**	−.07	−.11	.46**	.31*	.33*
20. T3 emotion FB	.48**	.38**	.26	.47**	.24	.26	.46**	.37**	.52**
21. T3 explain FB	.00	.11	.07	.15	−.25	.09	.33*	.28*	.37**
22. T3 penny	.00	−.09	−.02	.09	−.01	.04	.29*	.24	.28*
23. T3 Gordis	.42**	.52**	.27	.64**	−.04	−.09	.62**	.39**	.74**
24. T3 ms/hr	.18	.30*	.06	.29**	.07	−.21	.18	.29**	.24

Note. $N = 50$. T1 = Time 1; ToM = theory of mind; FB = false belief; Denham = Denham's (1986) study; MS/hr = number of mental-state
$^{*}p < .05$. $^{**}p < .01$.

= .51, .58, .34 for Times 1–2; Times 2–3; and Times 1–3, respectively, $p < .05$ for all. Correlations between adjacent time points remained significant even when individual differences in verbal and nonverbal ability at Time 1 were partialed out, $r(49)$ = .38 and .46 for Times 1–2 and Times 2–3, $p < .01$ for both. Correlations between each type of theory-of-mind task are also shown in Table 3, because the relationship between performance on different types of theory-of-mind task is of topical interest (e.g., Holmes, Black, & Miller, 1996). Early individual differences in both predicting and explaining false belief were correlated with each other and were relatively stable over the 13-month period (although, as expected, many of these correlations fell below significance once effects of verbal and nonverbal ability were taken into account). However, the deception tasks were not correlated with the other theory-of-mind tasks, and this could not be attributed to a floor effect (see Table 1). One possible account for this dissociation is that deception places much stronger demands than understanding false belief upon executive functions such as inhibitory control[2] (see Hughes, 1998).

Performance on the Denham (1986) task was also stable between Time 1 and Time 2, even when effects of both verbal and nonverbal ability were removed, $r(49) = .60$, $p < .001$. There was also good agreement between the different measures of emotion understanding used at each time point. In particular, with effects of verbal and nonverbal ability removed, children's scores on the mixed-emotion task were highly correlated with both the adequacy of their interview response 6 months earlier, $r(49) = .64$, $p < .001$, and with scores for affective perspective taking 13 months earlier, $r(49) = .50$, $p < .001$. Individual differences in children's emotion understanding were therefore robust across measures and stable over a 13-month time period.

Despite the logarithmic increase in the frequency of children's use of mental state terms, individual differences in mental-state talk were also stable. Correlations across adjacent time points remained significant even when effects of verbal and nonverbal ability at Time 1 were removed, $r(49) = .31$ and .30 for Time 1–2 and Time 2–3, $p < .05$. That is, within the specific context of dyadic play between friends, individual differences in frequency of mental-state talk appeared reliable across a 13-month period.

Relations Between Domains

Are Individual Differences in Understanding Mind and Emotion Associated?

The correlations for all three time points between performances on tests of theory of mind and emotion understanding (as well as rates of mental-state talk) are shown in Table 3. As the table indicates, children's performances were closely correlated both within and across tasks of understanding belief

[2] This proposal differs from Russell, Mauthner, Sharpe, and Tidswell's (1991) account, in that inhibitory control is predicted to be more strongly related with deception than with false-belief understanding, because only deception entails communication that runs directly counter to the Gricean truth maxim (see Hughes, 1998).

10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
—														
.58**	—													
.54**	.45**	—												
.35**	.23	.53**	—											
.27	.21	.33*	.20	—										
.37**	.43**	.43**	.38**	.32*	—									
.36**	.51**	.47**	.36**	.21	.46**	—								
.46**	.40**	.40**	.28**	.27	.47**	.28**	—							
.49**	.36**	.41**	.36**	.28*	.48**	.38**	.30*	—						
.31*	.32*	.36**	.27	−.13	.25	.22	.32	.50**	—					
.46**	.40**	.39**	.26	.24	.39**	.31*	.49**	.81**	.29*	—				
.37**	.24	.27	.18	.33*	.20	.29*	.30*	.61**	.12	.35**	—			
.26	.13	.20	.37**	.10	.24	.17	.24	.60**	.28*	.38**	.16	—		
.64**	.66**	.63**	.44**	.37**	.65**	.64**	.62**	.47**	.32*	.49**	.30*	.24	—	
.27	.28**	.24	−.12	.35**	.14	.10	.32*	.18	.11	.28*	.26	.04	.32*	—

terms per hour; T2 = Time 2; Cassidy = Cassidy et al.'s (1992) study; T3 = Time 3; Gordis = Gordis et al.'s (1989) study.

and emotion (although associations with performance on the deception tasks were less consistent). In general, these correlations remained significant even when individual differences in general ability were taken into account: All but 10 individual task correlations significant at the $p < .01$ level in Table 3 remained significant ($p < .05$) when effects related to verbal and nonverbal ability were removed. (Although most correlations were significant at all three time points, the relations between tasks appeared strongest at Time 2. This may reflect floor and ceiling effects on some tasks at Time 1 and Time 3.)

The interdependence between understanding belief and emotion is well illustrated by the emotion false-belief task, in which children were asked to predict both a story-character's false belief and the emotion provoked by this belief. Because this task was originally developed as a test of understanding belief, it was included within the theory-of-mind aggregate. However, it could be argued that the task also measures children's understanding of emotion. This dilemma was addressed empirically by using Fisher's z transforms to compare the strength of correlations between emotion false-belief performance and each aggregate score (first subtracting emotion false-belief scores from the theory-of-mind aggregate). This analysis showed that at Time 1, emotion false-belief task was significantly more strongly correlated with the remaining theory-of-mind tasks than with the emotion-understanding aggregate, $r(49) = .44$ and .16, $z = 2.13$, $p < .05$. At Times 2 and 3, this difference was nonsignificant, $r(49) = .46$ and .43; .56 and .49 at Times 2 and 3, respectively. Overall, then, performance on this task was at least as closely associated with understanding belief as with understanding emotion.

Close ties between understanding false belief and emotion were also apparent when the relation between aggregate scores in each domain was examined across time points. For example, a regression analysis showed that initial differences in emotion understanding predicted aggregate theory-of-mind scores 13 months later ($\beta = .54$, $R^2 = .29$), $F(1, 48) = 20.0$, $p < .001$. (Note that the standardized regression coefficient beta automatically controls for differences in range between measures.) A stepwise regression analysis was carried out to examine whether the predictive relation between early emotion understanding and later theory-of-mind performance was independent of age and ability. Age, verbal ability, nonverbal ability, and Time 1 theory-of-mind scores were entered at the first step of the regression, so that autocorrelation within theory-of-mind performance was controlled. Together these factors predicted 32% of the variance in Time 3 theory-of-mind scores ($R^2 = .32$), $F(4, 45) = 7.1$, $p < .001$. At the second step, Time 1 emotion-understanding scores independently predicted Time 3 theory-of-mind scores ($\beta = .32$, $p < .05$; $\Delta R^2 = .06$), $F(5, 44) = 5.4$, $p < .05$.

Separate regression analyses were conducted to examine the predictors of mixed-emotion understanding at Time 3. In a simple linear regression, initial theory-of-mind performance was a significant predictor of performance on the mixed emotion task (Gordis et al., 1989) 13-months later ($\beta = .42$; $R^2 = .18$), $F(1, 48) = 10.2$, $p < .01$. As before, a stepwise regression analysis was then carried out to examine whether this association was independent of age and ability, controlling for autocorrelations within emotion understanding. Age, verbal ability, nonverbal ability, and affective perspective taking were all entered at the first step and together predicted 48% of the variance in Time 3 Gordis scores, $F(4, 45) = 12.2$, $p < .001$, although only age ($\beta = .30$) and affective perspective-taking ($\beta = .37$) scores were significantly predictive of scores for understanding mixed emotion at Time 3. Adding Time 1 theory-of-mind scores at the second step did not increase the explained variance in Time 3 Gordis scores, indicating no independent effect of early differences in understanding belief in predicting later differences in understanding mixed emotions.

Relations Between Task Performances and Mental-State Talk

Even when effects of verbal and nonverbal ability were partialed out, individual differences in children's hourly rates of mental-state talk at Time 1 were significantly correlated ($p < .05$) with performance 13 months later in false-belief explanation, $r(49) = .31$; in predicting actions or emotions based on false belief, $r(49) = .31$ and .29, respectively; and in understanding mixed emotions, $r(49) = .50$. These associations were investigated using stepwise regression analyses, to control for effects of age, general ability, and autocorrelation effects.

The first analysis examined whether initial rates of mental-state talk predicted aggregate theory-of-mind scores at Time 3. When considered alone as a predictor of theory-of-mind performance, initial rates of mental-state talk accounted for 20% of the variance in Time 3 theory-of-mind scores ($\beta = .45$, $R^2 = .20$), $F(1, 48) = 12.0$, $p < .001$. Note that this relation was not strengthened when only "genuine" mental-state references were included ($\beta = .36$, $R^2 = .13$), $F(1, 48) = 7.0$, $p < .01$. Next, a stepwise regression analysis was conducted to examine whether the longitudinal association between individual differences in initial rates of mental-state talk and Time 3 theory-of-mind performance was independent of effects related to age, general ability, and autocorrelation with initial theory-of-mind performance. Age, verbal ability, nonverbal ability, and Time 1 theory-of-mind scores were entered at the first step of the regression, and together explained 32% of the variance in Time 3 theory-of-mind scores ($R^2 = .32$), $F(3, 46) = 7.1$, $p < .001$. At the second step, Time 1 frequency of mental-state talk remained marginally significantly associated with Time 3 theory-of-mind scores ($\beta = .25$, $\Delta R^2 = .05$), $F(5, 44) = 5.2$, $p = .07$. Note that the reverse relation, between initial theory-of-mind performance and Time 3 mental-state talk, was nonsignificant ($\beta = .18$), even before factors such as age and ability were taken into account.

The second analysis examined whether initial rates of mental-state talk predicted mixed-emotion understanding at Time 3. When mental-state talk was considered alone as a predictor of Time 3 mixed-emotion scores, a significant association was found ($\beta = .39$, $R^2 = .15$), $F(1, 48) = 8.8$, $p < .005$. However, a stepwise regression analysis with age, verbal ability, nonverbal ability, and Time 1 Denham scores (Denham, 1986) entered at the first step showed that these variables together explained 48% of the variance in Time 3 theory-of-mind scores ($R^2 = .48$), $F(4, 45) = 12.2$, $p < .001$; and Time 1 mental-state talk entered at the second step did not increase the explained variance, indicating that the association between initial mental-state talk and Time 3 emotion understanding was not independent of effects related to these factors. Note also from Table 3 that the correlation between Time 1 Denham performance and Time 3 mental-state talk was nonsignificant, and so this relationship was not examined.

Gender Differences

No effects of gender were found for any task measure at any time point (see *Developmental Change in Task Performances,* in the Results section). However, a repeated-measures ANOVA of (log transformed) mental-state talk (with standard vocabulary scores at Time 1 covaried) showed that girls displayed more frequent reference to mental states than boys, both per hour, $F(1, 48) = 4.6$, $p < .05$, and per speaker turn, $F(1, 48) = 6.6$, $p < .01$. Although the Gender × Time interaction terms were nonsignificant, differences between boys and girls was clearest at the final time point. Post hoc analyses (Tukey's honestly significant difference) at this time point showed that girls displayed more mental-state talk than boys per hour, $F(1, 48) = 7.9$, $p < .01$, and per speaker turn, $F(1, 48) = 12.2$, $p < .001$. Girls also used a greater variety of mental-state terms, $F(1, 48) = 5.8$, $p < .02$; made a greater number of genuine mental-state references per hour, $F(1, 48) = 5.6$, $p < .02$; and showed a greater number of mental-state references in the pragmatic context of shared interest, $F(1, 49) = 5.8$, $p < .02$. However, these findings should not be treated as separate effects, because the observational measures are highly intercorrelated, $r(49) < .75$, $p < .001$. Instead, these findings provide converging evidence that mental-state talk is both more frequent and developmentally more advanced in girls compared with boys.

Discussion

The main findings of the study can be summarized as follows. As expected, over the 13 months of the study, children showed marked improvements in their performance on tests of understanding false belief and affective perspective taking. Individual differences in these two domains were related to each other and stable over time. In addition, performance on both theory-of-mind and emotion understanding tasks was correlated with the frequency of children's mental-state talk in dyadic play with friends. In fact, initial individual differences in the frequency of mental-state talk in this social context were significantly associated with theory-of-mind performance more than a year later. Significant developmental change was also observed in the conversations recorded between friends. For example, children showed a steady logarithmic increase in their hourly rates of mental-state talk over the course of the study. Children also showed three important qualitative changes in their mental-state talk across time: (a) an increase in the proportion of genuine references to mental states; (b) a shift from referring primarily to their own mental states to referring at least as often to shared or others' mental states; and (c) a related shift away from mental-state talk as neutral commentary toward mental-state talk in the context of shared interest.

What are the implications of the above findings for the issues raised in the introduction? In their early longitudinal investigation of the relations among theory of mind, emotion understanding, and young children's interactions with others, J. Dunn, Brown, Slomkowski, et al. (1991) highlighted both the continuity in individual differences between children and the potential positive effects of child–child interactions (in the case of siblings) on children's growing mental awareness. The results of this study support each of these conclusions. Individual differences between children were stable over 13 months, both for task performances and for observational measures of children's conversational reference to mental states. The stability in task performance provides a reassuring counterpoint to recent suggestions that false-belief tasks have poor test–retest reliability (Mayes et al., 1996) and confirms the utility of the multitask aggregate approach used in the present study.

The stability of mental-state talk is also important, as it provides a first step toward exploring the significance of individual differences in young children's reference to mental states in their conversations with friends. Previous studies (e.g., Brown et al., 1996) have indicated significant developmental shifts in key conversational partners for mental-state talk. In this study, individual differences in frequency of children's mental-state talk with friends were highly stable, indicating developmental continuity in children's mental-state reference within specific conversational contexts. At the same time, as outlined above, the results of this study also indicate developmental changes in the quality of children's mental-state talk. However, these changes (e.g., increase in proportion of reference to others' mental states) may simply reflect baseline changes in children's interactions with friends over time. This possibility deserves further study.

The present results also extend findings from previous studies focused on siblings that indicate a positive effect of child–child interactions on children's developing social understanding (e.g., J. Dunn, Brown, Slomkowski, et al., 1991; Lewis et al., 1996; Perner et al., 1994). In this study, children who engaged in frequent reference to mental states in conversations with their friends not only showed higher concurrent false-belief understanding but also were more likely than their peers to perform well on false-belief tasks more than a year later. Remarkably, this relation remained marginally significant even when initial differences in theory-of-mind performance, age, verbal ability, and nonverbal ability were all taken into account.

Another topic raised in the introduction was the ecological validity of laboratory-based assessments of children's understanding of mind. The results of this study offer qualified support for these tasks. Children's performance on standard false-belief tasks was correlated with their mental-state talk at all three time points. However, once individual differences in verbal and nonverbal ability were taken into account, this relation fell below significance at two of the three time points. Moreover, although girls showed both more frequent and developmentally more advanced mental-state talk than boys, there were no gender differences in performance on the false-belief tasks. This suggests that although laboratory tasks provide a quick method of assessing children's social insight, they are less likely to prove especially sensitive to individual differences in how children apply their understanding of mind to the everyday social world.

Also raised in the introduction was the question of whether individual differences in children's understanding of emotions and beliefs are related. Unlike the findings reported by J. Dunn (1995), false-belief performance and emotion understanding were closely associated in this study. In fact, early affective perspective taking predicted theory-of-mind performance 13 months later, even when initial differences in theory of mind, age, and general ability were all taken into account. These results suggest close links between children's understanding of feelings and their understanding of belief. The discrepancy between the present findings and those reported by J. Dunn (1995) can be explained in at least two ways. One possible account is that the greater number of theory-of-mind tasks used in the present study provided a more robust aggregate than the single task measure used by J. Dunn (1995). A second possibility is that the differential sequelae of understanding mind and emotion emerge gradually over development; whereas J. Dunn (1995) compared task performance at 40 months with social adjustment 30 months later, the present study involved only a 13-month period. Some of the findings in the present study are consistent with this developmental account. For example, although both emotional perspective taking and false-belief comprehension were correlated with mental-state talk, once factors such as age, ability, and initial task performance were taken into account, longitudinal associations were more clear-cut between mental-state talk and later theory-of-mind performance rather than later emotion understanding. On balance, however, the results of the present study suggest general rather than specific relations between social understanding and conversational displays of mental-state awareness.

In contrast with the posited broad nature of relations between mental-state talk and sociocognitive ability, findings from previous studies suggest specific contrasts in mental-state talk across different conversational contexts (e.g., with different conversational partners, Brown et al., 1996; in pretend vs. nonpretend play, Hughes & Dunn, 1997). In addition, it has been proposed that such context effects show developmental change (Brown et al., 1996). In support of this view, in the present study, conversations were recorded between friends in the same play situation at all three time points, and yet the nature of children's mental-state talk changed significantly over time. For example, there was a striking increase in the proportion of mental-state references that occurred within the context of shared interest. The simplest interpretation of this finding is that children showed a baseline increase in their levels of shared interest at each time point. This possibility is currently being addressed through a study of developmental changes in the coordination of play between these young friends.

At the same time, context effects caution against drawing general conclusions about developmental changes in children's mental-state talk from any one social situation: Conversations between older friends or between children who are not close friends, or between children and their parents, may tell a different developmental story. In the final part of this article, we compare the present methods and findings with those reported by Bartsch and Wellman (1995) in their classic study of children's talk about the mind.

First, the present study was based on a smaller language sample than the Child Language Data Exchange System (CHILDES; MacWhinney & Snow, 1985) database used by Bartsch and Wellman's (1995) study (just under 20,000 utterances as compared with 200,000). Second, the present study involved children from a much narrower age range than was possible with the CHILDES database (47–60 months, as compared with 18–72 months in Bartsch and Wellman's study). Despite these limitations, the present study also carried a number of advantages over the CHILDES database. First, the combination of observational and experimental approaches adopted in the present study enabled direct empirical comparisons to be made between developmental changes in children's mental-state talk and their improving performance on theory-of-mind tasks. Second, transcription from videotape (as compared with audiotape) allowed more sensitive coding of the context of children's conversations. Third, 50 children were involved in this study (as opposed to only 10 in the CHILDES sample), improving the generalizability of the present findings and enabling issues of individual differences to be addressed.

Bartsch and Wellman (1995) focused on the onset of specific mental-state terms (rather than their frequency), and so adopted the conservative strategy of restricting analysis to cases of "genuine" mental-state reference (but included terms of desire as well as cognition). Previous researchers have found that mental-state terms typically appear in children's vocabularies in the 3rd year of life (Bretherton, McNew, & Beeghly-Smith, 1981). At the start of this study, the preschoolers were all in their 4th year (precluding calculation of onset ages), but because the dataset was more tractable in size than the CHILDES dataset, it was possible to calculate frequencies of mental-state reference. In addition, it has been argued that children's understanding of conversational uses of mental-state terms to modulate assertions emerges in parallel with (and may depend on) developments in theory of mind (Moore, Bryant, & Furrow, 1989), and so conversational uses of mental-state talk were included in the present study. In support of the claim made by Moore and his colleagues, "conversational" and "genuine" uses of mental-state talk were both associated with theory-of-mind task performance in the present study.

As can be seen from some of the methodological differences described above, the aims of the two studies should be seen as complementary. However, on at least one issue, the studies overlap and lead to different conclusions. This issue is the question of developmental changes in the referent of children's mental-state talk, a question that is pivotal to an ongoing debate between two contrasting perspectives on children's mentalistic development: theory-theory and simulation theory. The central claim of the theory-theory view is that children acquire an understanding of mind through a process of theory building (such that a valid analogy can be made between developmental changes in children's understanding of mind and conceptual shifts within scientific theories). This claim is based on the assumption that children do not have privileged self-knowledge of their own mental states but rather construct a theory of mental life that applies equally to self and other (Gopnik, 1993). In contrast, simulation theory holds that children begin by identifying their own mental states through introspection, and then generalize these mental states to other people through the imaginative process of "simulation"—putting oneself in the place of another so as to vicariously experience what that person might think or feel (Gordon, 1986; Harris, 1991).

According to Bartsch and Wellman (1995), analyses of the CHILDES database suggest no clear difference between the onset for children referring to their own versus others' mental states, and so support the theory-theory account of development in children's understanding of mind (rather than simulation theory). The present dataset concerns frequency of mental-state talk (rather than age of onset), and so have only tangential implications for this debate. Nevertheless, the children in this study showed a clear developmental shift, with a significant increase over time in the proportion of mental-state terms used to refer to others' (rather than their own) mental states. This shift is hard to explain within theory-theory but is consistent with simulation theory and with findings from another recent study of development in young children's mental-state talk (Imbens-Bailey, Prost, & Fabricius, 1997). However, before concluding that the findings from the two studies are contradictory in this respect, one should recall that all the conversations in the present study were recorded in the context of dyadic play between friends, whereas most of the utterances analyzed by Bartsch and Wellman were between children and adults. Conversational partners are known to exert significant influences on children's mental-state talk (Brown et al., 1996; Furrow et al., 1992; Hughes & Dunn, 1997) and may well explain the discrepancy between findings from the two studies. Alternatively, as Perner (1996) suggested, it is possible that a mix of both theory-building and simulation processes are involved in children's developing understanding of mind. The situation we studied, play between friends, through its demands for imaginary and cooperative interaction, may especially foster and support children's simulations of others' minds.

References

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding [Special issue]. *Cognition and Emotion, 9,* 151–165.

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

Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to belief and desires. *Child Development, 60,* 946–964.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* New York: Oxford University Press.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In M. Lamb & L. Sherrod (Eds.), *Social cognition in infancy* (pp. 333–373). Hillsdale, NJ: Erlbaum.

Brody, L. R. (1985). Gender differences in emotional development: A review of theories and research. *Journal of Personality, 53,* 102–149.

Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings and mothers. *Child Development, 67,* 836–849.

Brown, J. R., & Dunn, J. (1996). Continuities in emotion understanding from three to six years. *Child Development, 67,* 789–802.

Cassidy, J. C., Ross, D. P., Butkovsky, L., & Braungart, J. M. (1992). Family–peer connections: The roles of emotional expressiveness within the family and children's understanding of emotions. *Child Development, 63,* 603–618.

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development, 57,* 194–201.

Denham, S. A., McKinley, M., Conchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development, 61,* 1145–1152.

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition and Emotion, 9,* 187–201.

Dunn, J., Brown, J. R., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27,* 448–455.

Dunn, J., Brown, J. R., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62,* 1352–1366.

Dunn, J., & Hughes, C. (1998). Young children's understanding of emotions within close relationships *Cognition and Emotion 12, 2,* 171–190.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale.* Windsor, UK: NFER-Nelson.

Furrow, D., Moore, C., Davidge, J., & Chiasson, L. (1992). Mental terms in mothers' and children's speech: Similarities and relationships. *Journal of Child Language, 19,* 617–631.

Gopnik, A. (1993). How we know our minds: The illusions of first-person knowledge of intentionality. *Behaviour and Brain Sciences, 16,* 1–14.

Gordis, F. W., Rosen, A. B., Grand, S. (1989, April). *Young children's understanding of simultaneous conflicting emotions.* Poster presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1,* 158–171.

Gottman, J. M. (1986). The world of coordinated play: Same- and cross-sex friendship in young children. In J. M. Gottman, & J. G. Parker (Eds.), *Conversations of friends: Speculations on affective development* (pp. 139–191). Cambridge, UK: Cambridge University Press.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory-of-mind task performance of subjects with autism. *Child Development, 66,* 843–855.

Harris, P. L. (1991). The work of the imagination. In A. Whiten (Ed.), *Natural theories of mind* (pp. 283–304). Oxford, UK: Basil Blackwell.

Harris, P. L., Johnson, C., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3,* 379–400.

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67,* 1–13.

Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false-belief understanding in a Headstart population. *Journal of Experimental Child Psychology, 63,* 263–285.

Howes, C. (1988). Peer interaction of young children. *Monographs of the Society for Research in Child Development, 53* (Serial No. 217).

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16,* 233–253.

Hughes, C., & Dunn, J. (1997). "Pretend you didn't know": Young children's talk about mental states in the context of pretend play. *Cognitive Development, 12,* 477–499.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248.

Imbens-Bailey, A. L., Prost, J. H., & Fabricius, W. V. (1997). *Perception, desire and belief in me and you: Young children's reference to mental states in self and others.* Manuscript submitted for publication.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory-of-mind development in young children. *Developmental Psychology, 32,* 70–78.

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

Lewis, C., Freeman, N., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. M. (1996). The social basis of theory of mind: Influences of siblings and, more importantly, interactions with adult kin. *Child Development, 67,* 2930–2947.

MacWhinney, B., & Snow, C. E. (1985). The Child Language Data Exchange System. *Journal of Child Language, 12,* 271–295.

Mayes, L. C., Klin, A., Tercyak, K. P., Jr., Ciccetti, D. V., & Cohen, D. J. (1996). Test–retest reliability for false-belief tasks. *Journal of Child and Adolescent Psychiatry, 37,* 313–319.

McCarthy, D. (1972). *Manual for the McCarthy Scales for Children's Abilities.* New York: Psychological Corporation.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60,* 167–171.

Newcomb, A. F., & Bagwell, C. (1995). Children's friendship relations: A meta-analytic review. *Psychological Bulletin, 117,* 306–347.

Oswald, D. P., & Ollendick, T. (1989). Role taking and social competence in autism and mental retardation. *Journal of Autism and Developmental Disorders, 19,* 119–128.

Perner, J. (1996). Simulation as explication of predication-implicit knowledge about the mind: Arguments for a simulation–theory mix. In P. Carruthers & P. K. Smith (Eds.), *Theories of theory of mind* (pp. 90–104). Cambridge, UK: Cambridge University Press.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65,* 1228–1238.

Ruble, D. N., & Martin, C. L. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional and personality development* (5th ed., pp. 933–1016). New York: Wiley.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9,* 301–314.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to a mental state. *Cognition, 14,* 301–321.

Slomkowski, C., & Dunn, J. (1996). Young children's understanding of other people's beliefs and feelings and their connected communication with friends. *Developmental Psychology, 32,* 442–447.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33,* 591–605.

Taylor, M., & Carlson, S. (1997). The relation between individual differences in fantasy and theory of mind. *Child Development, 60,* 436–455.

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

Youngblade, L., & Dunn, J. (1995). Individual differences in young children's pretend play with mother and sibling: Links to relationships and understanding of other people's feelings and beliefs. *Child Development, 66,* 1472–1492.

Received July 1, 1997
Revision received January 25, 1998
Accepted January 25, 1998 ■