CHILD DEVELOPMENT

Child Development, July/August 2014, Volume 85, Number 4, Pages 1535–1551

The Contribution of Symbolic Skills to the Development of an Explicit Theory of Mind

Angeline S. Lillard
University of Virginia

Robert D. Kavanaugh
Williams College

Theorists have speculated about the symbolic underpinnings of theory of mind (ToM), but no study has examined them across the main developmental span of ToM. Here, the onset of symbolic understandings in three domains (pretend play, language, and understanding representations) and ToM was examined. Fifty-eight children were tested on batteries of tasks four times from ages 2.5 to 5 years. Some significant interrelations among variables were seen at each age level. Canonical correlation analysis found that a subset of the symbolic variables was significantly related to ToM at ages 4 and 5, providing the best evidence to date that ToM is undergirded by a symbolic element that also supports language, pretend play, and representational understanding.

A major achievement of early childhood is realizing that unobservable mental states, such as thoughts, beliefs, and emotions, underlie much of human behavior, an achievement referred to as a theory of mind (ToM; Premack & Woodruff, 1978). The importance of developing ToM, at least in Western societies, cannot be overestimated (Wellman, 2011). It is essential in communities that rely heavily on the exchange of information, ideas, and points of view through shared discourse. Developmentally, ToM represents a substantial change in the way that children engage with others. Put simply, a well-developed ToM is the key to realizing that behavior is motivated. People do not act without cause, but rather behave as they do because of what they think, know, feel, or believe.

Although we know a lot about ToM, such as what develops when (Wellman, 2011), that from ages 3 to 6 its latent factor structure is stable (Hughes, Ensor, & Marks, 2011), and that it predicts conceptually related factors like mental state talk, friendship interactions, and sociometric status (Capage & Watson, 2001; Hughes et al., 2011; Razza & Blair, 2009), much is still to be explained regarding how ToM develops. Some environmental features are important. For example, parents' attention to mental states appears to contribute to ToM (Meins et al., 2002; Ruffman, Perner, & Parkin, 1999), and having different-aged siblings or classmates is also often (but not always) associated with earlier development of ToM (Cassidy, Fineberg, Brown, & Perkins, 2005; Cutting & Dunn, 1999; Lewis, Freeman, Kyriakidou, Maridaki-Kassotaki, & Berridge, 1996; Perner, Ruffman, & Leekam, 1994; Ruffman, Perner, Naito, Parkin, & Clements, 1998; Wang & Su, 2009). Specific child characteristics are also associated with earlier development of ToM. Three child factors that have been repeatedly and concurrently correlated with ToM are children's level and frequency of pretend play (see Lillard et al., 2013), language ability (Astington & Baird, 2005; Hughes et al., 2005), and executive function (Carlson & Moses, 2001; Hughes, 1998). Two of these abilities—pretend play and language—are considered representational or symbolic. This study asks whether symbolic abilities used in pretending and language might actually undergird ToM. We are particularly concerned with an explicit ToM—revealed when a child expressly acknowledges that the world can be represented in multiple ways—rather than implicit ToM examined with looking time procedures (Apperly & Butterfill, 2009).

Both authors are affiliated with the Departments of Psychology.

We thank Joseph Lucia, Reka Daróczi, Leigh Owens, Mary Lindeke, Kyle Skor, and Elise Piazza for their assistance with data collection; John Nesselroade and Matt Lerner for statistical advice; and Paul Harris for comments on an earlier draft of this manuscript. This research was supported in part by NSF Grant 1024293, a Brady Education Foundation Grant, and a University of Virginia Sesqui award to ASL, and a Williams College Grant to RK.

Correspondence concerning this article should be addressed to Angeline S. Lillard, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904. Electronic mail may be sent to Lillard@virginia.edu.

© 2014 The Authors
Child Development © 2014 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2014/8504-0017
DOI: 10.1111/cdev.12227

Three well-known assessments of an explicit ToM are false belief (Wimmer & Perner, 1983), visual perspective taking (Flavell, Everett, Croft, & Flavell, 1981), and appearance–reality (Flavell, Green, & Flavell, 1986). These tasks, which have been used together in previous studies (e.g., Carlson, Mandell, & Williams, 2004), require that one indicate that a mental representation differs from the reality to which it refers—prime evidence for a ToM (Dennett, 1978). For example, a child must say that Maxi thinks his chocolate is in one place (his mental representation), although it is actually in another; that a given picture can look right side up to one person while simultaneously appearing upside down to another; and that although a wax object is actually a candle, it looks like (is visually represented as) an apple. In this way explicit ToM is symbolic at two levels: One must understand that mental stuff symbolizes real-world stuff (Perner, 1991), and one must use words or symbolic gestures to convey that understanding. Symbolic development, then, theoretically seems key to the acquisition of an explicit ToM.

The study reported here examines the relations between key symbolic *antecedents* of ToM at ages 2.5, 3, and 4, and ToM (as measured by explicit tasks) at ages 4 and 5. ToM was measured by the tasks just described, as well as use of mental state language when telling a story, and a limited-view perspective taking task (Chandler & Helm, 1984). Canonical correlation (Hotelling, 1935) was used to examine possible underlying relations across ToM and different symbolic tasks measuring development in three domains: pretense, language, and representation. To our knowledge this is the first study to examine this range of symbolic skills and their interrelations across the preschool period, from their emergence through the major development of an explicit ToM. Next, we discuss these three symbolic skills.

Potential Symbolic Antecedents of ToM

A symbol is something that is intentionally used to represent something other than itself (DeLoache, 2004). Examples are using the word *dog* to represent a dog, using a model to represent territory, and using a stick to represent a horse in pretend play. Piaget (1926, 1962) made clear the joint symbolic underpinnings of pretend play and language in discussions of the semiotic function, and Perner (1991) noted how, for symbolic reasons, both pretend play and language are conceptually related to ToM. Here, we first discuss the development of using symbols like pictures and scale models to represent a physical space, then move on to discuss pretend play and language.

Understanding Representations: Scale Models and Pictures

In the classic scale model task, a stuffed Snoopy is hidden (without the child's viewing) in a big room, and then a miniature Snoopy is (in full view of the child) hidden in the corresponding place in a doll house–sized replica room; children are asked to use the latter information to find the big Snoopy, who they are told is hiding in the exact same place in the big room (DeLoache, 2004). From 2.5 to 3 years of age children go from typically failing to find big Snoopy to reliably finding him. DeLoache attributed children's initial failure to an inability to handle dual representation: to see the scale model as both a thing in itself and a symbol for the larger room. As evidence of this, she showed that decreasing the salience of the model's identity as a little doll house by putting it behind glass improves task performance, whereas increasing its salience by having the child actually play with it prior to task administration impedes performance (DeLoache, Simcock, & Marzolf, 2004). Using a two-dimensional picture instead of the three-dimensional entity is another way to reduce salience, and indeed with pictures indicating the hiding place, children pass a version of this task a full 6 months earlier than with the three-dimensional scale model (DeLoache, 1991).

Interestingly, in all but one recent study, De Loache's scale model tasks have been administered in isolation, and no study to our knowledge has even examined the relation between the picture and scale model tasks to see whether they are associated (as would be expected if they tap a single underlying construct like symbolic development). The recent study just mentioned showed that scores on the scale model task were related concurrently to language (measured by the 3rd Test of Expressive Language [TELD–3]) at 30 months (r = .25), 36 months (r = .57), and 42 months (r = .70) of age (Walker & Murachver, 2012). Because language is a cardinal measure of symbolic activity, these relations are supportive of the idea that the scale model task is about symbolic understanding.

If these tasks and ToM tasks all tap symbolic activity, we would expect to see a relation between performance on picture and scale model tasks and ToM. Walker and Murachver (2012) did examine the relation between just the scale model task and false belief. Although not related at 30 months,

scale model performance at both 36 and 42 months predicted false belief performance 6 months later (at 42 and 48 months, respectively). In addition, at 36 months about one third and at 42 months half of the relation between scale model and false belief performance were attributable to language. This supports the idea that symbolic skills as assessed by the scale model task also undergird to some extent language and ToM, but the set of tasks and age ranges tested was limited.

In this study we gave children the point picture task at 2.5 years and the scale model task at 3 years of age, allowing us to see the relation between the tasks, and their current and predictive relation to other symbolic tasks, such as pretend play, language, and ToM.

Language

Language is the quintessential symbolic activity. Many studies show language is positively related to ToM (see Astington & Baird, 2005). Such studies have used a variety of measures of language, including measures of comprehension and production, and of vocabulary (also used as a proxy for IQ) and syntax. Relations to each type of measure are robust. Furthermore, the relation is predictive, with 2-year-old language, accounting for 45% of the variance in 4-year-old ToM (Watson, Painter, & Bornstein, 2001). The relation between ToM and language could at least in part be due to symbolic underpinnings of both.

A second potential reason for a language–ToM relation is syntactic (Harris, de Rosnay, & Pons, 2005), theoretically involving cognitive architecture supporting *sentential complements* structures (de Villiers, 2007). Sentential complements are subordinate clauses formed from verbs that convey information about knowledge and beliefs (e.g., *know, think*) and communication (e.g., *say, tell*). For example, "John believes that Jim is sleeping upstairs" contains a proposition (Jim is sleeping upstairs) and the statement's truth depends on the accuracy of an epistemic state (John's belief) rather than the actual state of affairs. Three- and 4-year-olds' memory for sentential complements accounts for a substantial amount of the variance in false belief understanding, with only a small portion of variance accounted for when one examines the reverse relation from false belief to sentential complements (de Villiers & de Villiers, 2000). As further evidence of a directional relation, after training on sentential complements, children's ToM scores increased (Lohmann & Tomasello, 2003).

In this study we assessed language at the first three time points using three different measures, each chosen for its prevalence in the literature at each age level, either overall or for use in prominent ToM studies involving children of that age level. At 2.5 years, children's parents filled out the MacArthur-Bates Communicative Development Inventory (MB–CDI; Fenson, Dale, Reznick, & Bates, 1994) measuring vocabulary and complexity of syntactic constructions. At 3 years, children were given the Peabody Picture Vocabulary Test (PPVT–III; Dunn & Dunn, 1997), and at 4 they were given the sentence comprehension subtest of the Clinical Evaluation of Language Fundamentals (CELF; Wiig, Secord, & Semel, 1992), and the sentential complements task (de Villiers & Pyers, 2002).

Pretend Play

Pretend play occurs when a child treats one situation or object as if it were a different one, in a spirit of fun. In so doing, the child maps his or her mental representation of the pretend entity or place on to the real (Lillard, 1993). There are at least two ways in which pretend play could contribute to ToM: through symbolic substitution that occurs just with objects, or "object substitution play," and additionally through "role play," in which the child practices with social content (Sachet & Mottweiler, 2013). For example, in pretending to be a baby, a child might consider the world from a baby's point of view, and therefore (theoretically) be better able to take a baby's perspective later when no longer pretending. Children's earliest pretending tends to involve object substitution without social content; role play emerges later.

Object substitution. Even the very earliest instances of object substitution pretend play involve symbols, as Piaget (1962) documented when his daughter at 15 months used the edges of a frayed cloth and the collar of her mother's coat to symbolize a pillow. Symbolic development in pretend production accrues in increasingly complex steps (McCune, 1995). Children initially engage in simple self- and other-directed acts, like pretending to feed themselves or another person with a spoon. They later advance to more clearly symbolic acts, like using a stick to represent that spoon. Although early pretending is often introduced by parents (Lillard, 2011), most children begin to pretend that one thing is another by the time they are 2 years old (Haight & Miller, 1993).

A key concern with basic object substitution pretend play is the relation between production and

comprehension of pretense. Leslie (1987) claimed that pretense production and comprehension must appear simultaneously in development. However, empirical tests suggest that, opposite to language, in laboratory tests pretend play production precedes comprehension (Hopkins, Smith, & Lillard, 2014). Harris and Kavanaugh (1993) showed that by 2.5 years, many children can follow and interpret an experimenter's pretend acts (see also Ma & Lillard, 2006), but production of pretend acts occurs much earlier.

Both pretend production and comprehension rely on symbols, but in the former case the child creates the symbol, whereas in the latter the child reads the symbols of others (and in testing, these others are typically strangers in a laboratory). Using a pretend symbol that one makes up might be easier than understanding a symbol made by a stranger, but interestingly, the relation between pretend production and comprehension has received little attention. Here, we administered tasks assessing pretend production (in free play with the mother) and comprehension (using variants of the Harris & Kavanaugh, 1993, tasks) at 2.5 years of age, so we could see their interrelation as well as their relation to later pretend play and other symbolic tasks. We know of no correlational studies linking very early pretend play to later ToM, but laboratory studies of object substitution and its concurrent relation to ToM in the preschool years have yielded mixed results (Lillard et al., 2013).

Role play. The second form of pretend play that could engender ToM is role play. Children begin role play pretense in the 2nd year, when they attribute agency and simple mental states (desire, perception) to dolls (Wolf, Rygh, & Altshuler, 1984). By 4 years, most children engage in role play by pretending to be other people (Harris, 2000), and about half of children even invent others out of thin air, creating imaginary companions (ICs); a special case of ICs occurs when children repeatedly animate a stuffed animal or other item, creating a personified object (Taylor, 1999).

All these forms of role play could contribute to ToM for several reasons. One is that children represent others' mental states when they engage in role play. In doing so, they simulate others' mental states, and simulation might be how typically developing people understand others' mental states (Harris, 1995). Even were simulation not involved, such role play could still help ToM by giving children practice in manipulating mental state concepts because pretend play often hinges on the attribution of mental states and emotions (fear, anger, and so on) that can arise and fall away in the changing contexts of pretense events (Dunn & Hughes, 2001). Both simulation and practice at representing mental states are often considered important in their own rights, but here we consider each with regard to its symbolic aspects: Both involve the quintessentially symbolic act of representing mental representations.

Many studies have reported correlations between social pretend or role play, including play with ICs, and ToM (see Lillard et al., 2013). However, the source and direction of the relation is not clear, so studies of the relations between pretend role play and ToM over time are of considerable interest.

Here, we examined role play at ages 3 and 4, both with tasks designed to induce it (having children complete a narrative with small figures, giving children a doctor's kit to play with, and asking them to telephone a pretend friend) and by asking them and their mothers about their ICs. We also used a less direct parent-completed measure of fantasy orientation that has been associated with role play and ToM in other studies (Taylor & Carlson, 1997).

Summary

This study examined three types of symbolic antecedents to ToM, namely, physical representations of space, language, and pretend play, in children 2–5 years of age. The main goal was to examine whether highly related underlying constructs could be derived across the sets of symbolic and ToM tasks. If such constructs were found, it would support the possibility that related symbolic competencies undergird both sets of tasks.

Method

Participants

Seventy-seven children (39 boys) and their mothers were recruited for the first wave; 77% returned in later waves, making a final sample of 58 (29 boys; M = 28.43 months, SD = 1.58, 24–31 months at Wave 1). Aside from one nonreturning child with substantial developmental delay, the nonreturners did not significantly differ from those who returned on any of the Wave 1 variables, so they are not discussed further. The Wave 2, 3, and 4 visits were conducted at 36, 48, and 60 months, respectively, ±2.5 weeks on either side of the child's birthday. Participants were from rural Massachusetts and were recruited through ads and flyers in local papers and day-care centers. They came primarily

from two-parent Caucasian homes (88%, or $n = 51$). About half of mothers ($n = 28$) had 2+ years of postgraduate education, 28% ($n = 16$) completed college only, 14% ($n = 8$) had 1–3 years of college, and the remaining 10% ($n = 6$) had no more than a high school education.

Procedure

All assessments took place in a 4.3 × 5 m laboratory playroom with a one-way mirror and appropriate home furnishings. Sessions were videotaped through the one-way mirror and also with a wide-angle camera high on the wall in the playroom. Different measures were used at each time point, as detailed later in this article. There was some missing data due to very occasional child inattention or noncompliance; two children also missed the Wave 3 visit.

Wave 1

At 24–31 months, children completed three tasks in a fixed order: pretense production in free play, pretense understanding, and symbolic understanding; their mothers completed the MB–CDI at home and returned it within 1 week of the test session.

Pretense production. Pretend production was measured in a 15-min free play session in the laboratory playroom, outfitted with a set of gender-neutral toys. Mothers were instructed to play with their child as they would at home. For the last 5 min of the session, a tea party set was added and participants were encouraged to play with it. Children's pretend play was scored using McCune's (1995) coding scheme, which differentiates three types of pretense: (a) *self-directed* pretend actions, such as sipping from an empty cup or simulating sleeping on a toy pillow; (b) *other-directed* pretense involving the mother or a toy doll, such as "pouring" tea into the mother's cup or "feeding" the doll from an empty toy bottle; and (c) use of object substitutes or imaginary objects, such as stating and acting as if a banana is a saxophone or "eating" a cookie without the use of a toy prop. Each event was coded to reflect the highest level to which it corresponded. A pretend episode began when children either picked up a toy or, with a toy already in hand, demonstrated one of the three levels. It ended when children discarded the object(s), began a new pretend action, discontinued the current pretend action for 10 s or more, or shifted to manipulative or sensorimotor play. Actions physically guided by the mother (e.g., taking the child's hand and "pouring tea" into a cup) were not scored. Each pretend act, regardless of level, was given 1 point, and these were summed for a child's pretend production score.

Coders watched the videotapes and wrote a description of each child pretend act that they observed, noting the time that it began, the duration, and the type. One coder scored all 58 tapes. One of two other reliability coders (trained college undergraduates or the first author) scored half of the final sample, randomly selected, for this and all other reliability coding in the study except where noted. Agreement criteria for this task were that the principal and reliability coder agreed (within 5 s) on both the time that a pretend action occurred and the type of pretense. The intraclass correlations for the two raters' pretend production scores were .97, .96, and .98 for Levels 1, 2, and 3, respectively. For all coding in this study, disagreements were resolved by discussion, with final decisions made by the second author.

Pretense comprehension. The pretense comprehension task (Walker-Andrews & Harris, 1993) involved both simple and complex transformations performed with a variety of everyday toys. Simple transformations involved two identical objects (e.g., two toy dogs); a pretend transformation was enacted on one (pretending to give one dog a bath), and the child was asked to respond ("Can you dry the dog that is all wet?"). Because the simple tasks could be solved by merely acting on the object on which the experimenter had just acted, complex transformations were also used. Complex transformations involved showing children two identical objects (two toy dogs), pretending to transform *both* of them (e.g., by pretending to give both dogs a bath, thus making them both "wet"), reversing or undoing the transformation on one of the two objects ("drying" one dog with a cloth), and then asking children to respond in a manner that would indicate that the pretend transformations were tracked ("Can you dry the dog that is all wet?"). Children saw three simple and three complex episodes in counterbalanced order. There were no differences in performance across the types so the two types were summed, resulting in pretense comprehension scores ranging from 0 to 6.

Point picture task. Following DeLoache (1991), children were asked to use a color photograph (20 × 25 cm), displayed on a table in a small anteroom, to determine the location of a doll hidden under a piece of furniture that corresponded to the photograph. Four trials were given, each involving

a distinct hiding location. Scores ranged from 0 to 4.

Language. Mothers filled out the MB–CDI (Fenson, Marchman, Thal, Dale, & Reznick, 2007), scored for the number of words their child knew (language production), and the number of correctly used syntactic constructions (language complexity).

Wave 2

When the children were within 2.5 weeks of their third birthday, they completed a play narrative (Wolf et al., 1984), the scale model task (De Loache, 1987), the PPVT–III (Dunn & Dunn, 1997), and an IC interview (Taylor & Carlson, 1997). Their mothers also did the IC interview and filled out a questionnaire pertaining to the child's fantasy orientation.

Play narrative completion task. After 5 min of warm-up play with the experimenter, children watched the experimenter enact with miniatures the beginnings of two narratives (in counterbalanced order): going to the zoo and having a picnic. The experimenter prompted children to complete the story by saying, "Can you show and tell me what happens next?" If children did not respond, the experimenter prompted them by asking, "What happens next?" Children's responses were scored using a 5-point system: 0 = *no response or simple manipulation of the props*, 1 = *response has a theme that is unrelated to the story stem and no evidence of active agency*, 2 = *the theme is related to the story stem but no evidence of active agency*, 3 = *the theme is not related to the story stem but shows evidence of active agency*, 4 = *the theme is related to the story stem and shows evidence of active agency*. The reliability coder independently scored a randomly selected 20% of children's narratives, and Cohen's kappa was .86.

Scale model. The scale model task followed the procedure in DeLoache's (1991) "hide-model" experiment. A scale model of the playroom (one seventh in size) was in the laboratory anteroom. The four playroom hiding locations were the same as those used in the point picture task given at Wave 1. Scores ranged from 0 to 4.

PPVT–III. The PPVT–III was administered according to the manual. Raw scores were used because there was little variation in the children's ages.

IC interview. Following the procedure in Taylor and Carlson (1997), children were told, "Now I'm going to ask you some questions about friends. Some friends are real like the kids who live on your street, the ones you play with. And some friends are pretend friends. Pretend friends are ones that are make-believe, that you pretend are real. Do you have a pretend friend?" If the child answered, "Yes," he or she was asked several more questions, for example, about the friend's name, gender, age, and appearance. Mothers were interviewed separately regarding whether their child had an IC, with the same follow-up questions as were directed at the child, if the mother said the child did have an IC. Child and maternal responses were examined by two coders who then judged whether the child had an IC (including a personified object, like a stuffed bear friend). Children were credited with an IC if either the child's or the mother's response indicated that the child truly had an IC. A second coder coded 100% of the sample and the intraclass correlation for IC status was .92.

Fantasy. Mothers were asked to write down children's favorite toys and play activities, and responses were scored for fantasy orientation. A score of 0 indicated the toy or activity was oriented to reality (e.g., puzzles) and a 1 indicated a fantasy orientation (e.g., play Ninja Turtles). Interrater agreement was calculated for a randomly selected 20% of the sample, and Cohen's kappa was .86.

Wave 3

When the children were 48 months, their pretend role play, language including sentential complements, ToM, and ICs were assessed. At this wave the intraclass correlation for IC status reliability coding was 1.00.

Role play. Two tasks were used to assess role play: a doctor kit task and a phone task.

Doctor kit task: Children were given a doctor kit with which to play with their mother for 7.5 min. Sessions were videotaped and transcribed to yield protocols that contained the verbal statements and actions of both the child and the mother. Children's statements were divided into "turns," defined as one person's utterances bounded by another person's, and then assigned to four categories, developed from prior coding schemes (Astington & Jenkins, 1995; Hughes & Dunn, 1997): *role statements*, in which children made a statement consistent with the role of either doctor or patient (e.g., "You need a shot"); *role assignments*, in which children explicitly assigned a role to either self or mother ("I'm the doctor now"); *third-party references*, in which children referred or spoke to an imaginary third person in the role of doctor or patient ("Nurse, you need to come down here quickly");

and *mental state references* ("I think I'll need that"). Mental state references and other categories could co-occur; for example, "I think I'll give you a shot now" would be both a mental state reference and a role statement. Children received a score for each category, plus an overall role play score summing the four categories. The intraclass correlation was .98.

Phone task: This task was based on Taylor's (2005) role play coding scheme. Children were shown a toy telephone, asked to name a friend whom they could pretend to call, and then encouraged to "call" the friend. One point was awarded for each of the following child actions: dialing, holding the phone receiver to the ear or mouth, talking on the phone, and listening on the phone. An additional point was given for each additional speaker turn (after the initial point for talking on the phone), defined as each time a child spoke; a pause started a new turn. The intraclass correlation was .97.

ToM battery. The Wave 3 ToM battery consisted of Level 1 and 2 perspective taking tasks from Flavell et al. (1981), an appearance–reality task from Flavell et al. (1986), and a standard location change false belief measure based on Wimmer and Perner (1983). For the Level 1 task, children were shown a paper depicting a cat on one side and a dog on the other. After asking a child to identify each animal, the experimenter held the card up vertically between himself and the child, so that the child could see only one side, and asked, "When you look at this, what do you see with your eyes?" and "What do I see with my eyes?" Children passed if they responded correctly to both questions.

For Level 2 perspective taking, children were shown a paper depicting a car. The paper was placed face up on the table between the experimenter and the child, and the child was asked whether the car looked right side up or upside down to him or her. The experimenter then rotated the card 180 degrees and asked again whether the car looked right side up or upside down to the child. Finally, without moving the card again, the experimenter asked, "And how does it look to me? Is it upside-down or right-side-up for me?" Children had to respond correctly to all three questions to receive credit for Level 2 understanding.

For the appearance–reality task, children were shown a sponge that was painted so it looked like a rock. After confirming that the object looked like a rock, children were told to pick up and feel the sponge, and the experimenter asserted that it really was a sponge. The experimenter then placed the sponge back on the table and asked, "When you look at this with your eyes right now, what does it look like, a rock or a sponge?" and then, "And what is it really and truly? Is it really, really a sponge, or is it really, really a rock?" One point was given for correctly answering both questions.

For the location change task, the experimenter explained (with appropriate props) that Miss Piggy and Kermit were playing together outside, but that Kermit had to leave, and he wanted to keep his favorite ball safe. Kermit put his ball into a red barrel, and then walked behind a wall, after which Miss Piggy moved the ball from the red barrel to the white cup. As memory checks, children were asked (a) where Kermit had left his ball and (b) where the ball was now. Then Kermit returned and children were asked the test question, "Where will Kermit look for his ball?" Children were given 1 point for correctly answering all three questions (two memory checks and the test question).

Because children received 1 point for each task on which they answered correctly, scores ranged from 0 to 4. The conceptual relations between these tasks are indicated by prior research showing good correlations when using more than one of each type of measure (Carlson et al., 2004; Frye, Zelazo, & Palfai, 1995; Gopnik & Astington, 1988; Taylor & Carlson, 1997). As we had just one binary measure for both location change and appearance–reality, and children were at a transitional age, examining intertask correlations here is unwarranted.

Language. Two language measures were used: the CELF and Sentential Complements.

CELF: The 10 test sentences comprising the Sentence Structure section of the *Stimulus Manual 1* were used; each sentence was accompanied by sets of three pictures. Children were asked to point to the picture showing the action described in each sentence (e.g., a picture of a boy pushing a girl accompanying the sentence, "The girl is being pushed by the boy"). Scores ranged from 0 to 10.

Sentential complements: Children's understanding of sentential complements (embedded clauses) was assessed with a task modeled on that of de Villiers (de Villiers & Pyers, 2002). Children were shown a booklet in which every page contained two pictures: a wide view of a scene with two people and a close-up showing part of the scene. For each picture, children heard a sentence containing an embedded clause describing what a protagonist said or thought about the scenario (e.g., "He told the girl there was a bug in her hair, but it was really a leaf," with the wide view showing a girl with something in her hair, and the close-up

showing a leaf in her hair). The experimenter asked, "What did he tell the girl?" There were four such tasks, two about what the protagonist said and two about what the protagonist thought; thus, scores ranged from 0 to 4.

Wave 4

At the final wave, at 5 years, children narrated a wordless storybook, and took several standard ToM tests and the Test of Emotion Comprehension (TEC).

Frog story. The story *Frog Where Are You?* (Mayer, 1969) is a 29-page wordless illustrated book about a boy on a search for his runaway frog. Children, who had been familiarized with the book earlier, were asked to go through the book and tell the story. Their narratives were coded for mental state terms (words referring to perceptions, desires, beliefs, intentions, emotions, etc.). Two coders scored all protocols, and the intraclass correlation for mental state references was .98.

ToM tests. ToM tasks at Wave 4 included variants of the location change (using different Sesame Street characters and props; in addition to asking where the character would look, children were also asked if he knew, for a 0–2 score) and two appearance–reality tasks (a rock–eraser and a pear–candle, for a 0–2 score), a false contents task, and a limited view task. The false contents task was given in the standard way (Perner, Leekman, & Wimmer, 1987) with a Band-Aid box containing crayons. The experimenter asked a preliminary question ("What do you think is inside this box?") and two test questions, namely, what did an uninformed puppet think was in the box, and what did the child think was in the box before the contents were revealed. All children were correct on the preliminary question at first pass. Scores on the two test questions were summed, so total scores on this item ranged from 0 to 2. For the limited view task, children were shown pictures of three animals, which were then covered so only a nondescript part of the animal was visible (Taylor & Carlson, 1997), and children were asked if a puppet who saw only the uncovered portion could identify the animal. Limited view scores ranged from 0 to 3. All scores were summed, so total ToM scores ranged from 0 to 9.

The Test of Emotion Comprehension. The TEC (Pons & Harris, 2000) measured children's understanding of mental states and emotions by having them listen to vignettes about a protagonist experiencing emotions ranging from simple (e.g., happy or angry) to complex (e.g., hidden emotions and mixed emotions). Children responded to each vignette by pointing to one of four facial displays that illustrated how the protagonist felt. Scores could range from 0 to 9.

Results

First, we present descriptive statistics and show interrelations among tasks within each wave. Next, we reduce the variable set and then report results of the canonical correlation analyses addressing whether a single common construct can be found across the pretend, language, and symbol tasks that is highly related to a construct underlying performance on the ToM tasks. Cross-lagged correlations among key tasks at different waves are reviewed there.

Descriptive Data and Within-Wave Correlations

Frequency distributions were created for each variable, and only two were not normally distributed (IC because almost half the sample had 0 and another half had just 1, but a few had 2–4, and sentential complements, which had a U-shaped distribution). Table 1 shows the means and standard deviations for the normally distributed variables at all waves, as well as the intercorrelations for all variables at that wave.

Beginning with Wave 1, the point picture task was related to both measures of language (vocabulary and complexity) and pretense comprehension, and showed a nonsignificant (r = .20) relation to pretense production. The two language subtests were strongly correlated, and vocabulary was also related to pretense comprehension. The lack of relation between pretense production (producing one's own symbol) and pretense comprehension is notable. For Wave 1 only, because of increased age variability, age-partialed correlations are given (in parentheses). As can be seen, correlations are somewhat attenuated when age was partialed out, but the same pattern of relations was largely obtained with the vocabulary, language complexity, and point picture task relations dropping below significance because of significant relations between age and all three of these variables (rs = .35, .28, and .39, respectively).

At Wave 2 the only significant correlation was between the model room task and the PPVT (r = .31). This echoes the relation between the symbol task and language at Wave 1, although here all

Table 1
Means, Standard Deviations, and Correlations at Waves 1 (Age 2.5), 2 (Age 3), 3 (Age 4), and 4 (Age 5), With Age-Partialed Correlations at Wave 1 in Parentheses

Tasks			Correlations			
Wave 1	*M*	*SD*	Vocabulary	Language complexity	Pretense production	Pretense comprehension
Point picture	1.79	1.62	.30* (.19)	.28* (.20)	.20 (.16)	.44** (.40**)
Vocabulary	501.47	131.13	—	.78** (.77**)	.06 (.02)	.36** (.32*)
Language complexity	23.00	11.00	—	—	.11 (.08)	.22† (.17)
Pretense production	25.03	10.64	—	—	—	−.06 (−.09)
Pretense comprehension	3.05	1.86	—	—	—	—

Wave 2	*M*	*SD*	PPVT	Play narrative	Fantasy	IC
Model room	1.62	1.31	.31*	−.08	.17	.16
PPVT	45.64	12.34	—	.10	.14	.18
Play narrative	2.14	2.12	—	—	.11	.24†
Fantasy	1.05	.99	—	—	—	.22†

Wave 3	*M*	*SD*	Sent. comp.	Role play	Telephone	ToM	IC
CELF	7.89	1.26	.19	.11	.02	.12	.03
Sent. comp.	(see text)	—	—	.31*	.02	.51**	.18
Role play	11.59	8.87	—	—	.46**	.33**	.12
Phone	2.48	1.88	—	—	—	.04	.21
ToM	2.66	1.05	—	—	—	—	.28*

Wave 4	*M*	*SD*	TEC	Frog MS
ToM	5.76	2.48	.37**	.13
TEC	4.91	1.45	—	.17
Frog MS	12.40	5.33	—	—

Note. PPVT = Peabody Picture Vocabulary Test; IC = imaginary companion; CELF = Clinical Evaluation of Language Fundamentals; ToM = theory of mind; TEC = Test of Emotion Comprehension; Frog MS = mental state terms in child's retelling of frog story. Age-partialed correlations are given in parentheses for Wave 1. †$p < .10$. *$p < .05$. **$p < .01$. (two-tailed)

children were within 2 weeks of their third birthday so age is less relevant. The pretend variables included at this wave, play narrative and fantasy orientation, were unrelated to the concurrent language and symbol variables. At Wave 2, 48 children (83%) did not have an IC or impersonate others regularly; 10 did so (17%), including 1 girl who had a score of 2 because she had two imaginary friends; there were trends for IC to be related to play narrative and fantasy orientation.

At Wave 3, sentential complements had a U-shaped distribution, with 23 children (40%) scoring 4, 13 (22%) scoring 0, and the remaining 17 fairly evenly spread across scores ranging from 1 to 3. It was rescored as a binary variable with scores that ranged from 0 to 2 recoded as 0 ($n = 25$) and scores of 3 and 4 recoded as 1 ($n = 30$). As shown in the table, at Wave 3 role play was significantly related to sentential complements, the phone task, and ToM. Consistent with other research, ToM was very highly related to sentential complements. IC was also related to ToM. Most children lacked an IC at Wave 3; 9 had one and 1 had two.

At Wave 4, children's emotion and ToM knowledge were significantly related. This makes sense, as several items on the TEC presuppose understanding mental states. However, neither was related to the number of mental state references in children's narratives of *Frog Where Are You?*, suggesting that spontaneously using ToM terms to describe a character in a (or at least this) narrative is unrelated to responding correctly to questions about mental states.

Data Reduction

The raw correlations, coupled with conceptual relations, suggested several variables should be

combined for the canonical correlation analyses. The first three noted below are within wave.

Language

The MB–CDI at Wave 1 had two subtests: production asking mothers about basic vocabulary and complexity asking about syntactic constructions. Both address early language proficiency and they were strongly related, even with age controlled, as was shown in Table 1. The scores were converted into *z* scores and summed to create a single MB–CDI measure.

Pretend Play

At Wave 3, both the telephone and the doctor role play tasks were aimed at the single construct of role play ability. As can be seen in Table 1, they were strongly related, so they also were converted into *z* scores and summed to create role play.

W4 ToM

At Wave 4, ToM was significantly correlated with the TEC, as shown in Table 1. Because both address ToM at age 5 (some TEC questions even concern false belief), they were combined by summing the two raw scores.

In addition to the three new within-wave variables just mentioned, three other variables were created by combining variables that had strong conceptual underpinnings but were measured at adjacent waves.

Symbol

The point picture task at Wave 1 and the scale model task at Wave 2 are believed to address the same underlying construct of symbolic understanding. (DeLoache, 1991). In keeping with this, they were strongly related, $r(56) = .45$, $p < .001$. The two 0–4 scores were added to create a single variable symbol.

Imaginary Companions

IC, which was measured at 36 and 48 months, was also reduced because the important element for our purposes is having had an IC at some time prior to or concurrent with the measure of ToM. Thus, we constructed a new variable IC. Across the first two time points, 32 children scored 0 because they did not have an IC at either time point, and 26 scored 1 because they had one or more ICs at either or both time points.

Language

The MB–CDI, the PPVT, and the CELF were age-appropriate tests assessing language at each wave. Clearly, these all have strong conceptual relations, and thus the three scores were combined (by summing *z* scores, with a new *z* score created from the two summed MB–CDI *z* scores) to create a single language score. Because sentential complements bears a strong conceptual relation to false belief *and* language, its inclusion could confuse a straightforward measure of the relation between language and ToM. It was therefore accorded a separate analysis.

Canonical Correlation

Overview of Canonical Correlation Technique

Canonical correlation (CC) identifies the existence of underlying commonalities among variables (Hotelling, 1935). This technique is in the family of generalized linear models, and can be viewed as an omnibus test for correlations. CC analysis takes two sets of variables and locates the underlying vector that best fits each set of variables and correlates most strongly with the vector underlying the paired set. It generates both a loading for each variable on the latent vector (like the factor loading in factor analysis), reflecting the shared variance between a variable and the vector, and a coefficient or weight for each variable in constructing the underlying vector. As with multiple regression, the weights (coefficients) can be unstable because they depend on which other variables are in the set, so more attention is typically given to the loadings. CC is the preferred data analytic method for this data set (rather than regression) for two reasons. First, we were predicting outcomes at two time points and used different sets of interrelated measures at each time point. Second, regression tests whether the variables add independently to the variance explained, whereas we were interested in how the overlapping ("symbolic") contribution of the three sets of variables (language, pretend, and symbolic understanding) predicts a construct underlying ToM. The canonical correlation syntax script supplied with SPSS 21 for Mac was used to run these analyses.

Canonical Analyses

For the main canonical analyses, the pretend variables mentioned above (pretend production and comprehension at Wave 1, fantasy orientation and play narrative at Wave 2, the combined role play variable at Wave 3, and IC at Waves 2 and 3), the combined language variable, and the combined symbol variable constituted Set 1. ToM at Wave 3, ToM at Wave 4, and mental state references in the frog story at Wave 4 constituted Set 2. This model produced a significant CC between the first vector (variate) underlying each set of variables, $R = .62$, Wilks's $\lambda = .40$, $\chi^2(24) = 41.42$, $p = .02$. Two additional variates were not significant (*p*s =.13 and .36, respectively), so will not be discussed. Table 2 shows the loadings, coefficients, and correlations of the variables. Loadings at or near .30 are considered notable. Play narrative is only near notable at .29, but is reasonably well weighted at .22. Particularly high loadings were obtained for symbol, language, role play, and IC, with pretend production loading less but still respectably. Several of these variables appear to involve symbolic generativity. The ToM measures that loaded highly on the related vector were W3 ToM and W4 ToM; using mental state language in one's narrative was unrelated. Pretend comprehension and fantasy have below threshold loadings on the Set 1 variate. ToM at age 5 (Wave 4) was more strongly related to the underlying vector than ToM at age 4, perhaps because of increased variance, and/or perhaps because the Level 1 perspective taking task is less important in its relation to symbolic understanding (it is a test of "cognitive connections"; Flavell, Green, & Flavell, 1990).

Note that 8 of 24 possible correlations between the three ToM measures and the other variables were significant (two-tailed), and another 5 showed a trend. The significant correlation between the play narrative and role play scores could derive from the fact that high scores on the play narrative task were given for taking on the perspective of the characters in the narrative (e.g., speaking for them). This would also explain its relation to ToM at Wave 4, although the lack of significant relation between play narrative and ToM at Wave 3 remains unclear; again, it could be due to restricted range as fewer scores contributed to ToM at Wave 3, or due to the inclusion of a Level 1 PT task.

Two additional analyses explored the degree to which sentential complements works as a language or a ToM measure by examining just the language and ToM constructs. When sentential complements was added to the predictor language variables, the first of two canonical variates was significant, $R = .66$, Wilks's $\lambda = .54$, $\chi^2(8) = 30.7$, $p < .001$. The complements task had a loading of .88; of concern is that it also had a very high cross-loading on the ToM variate, .58, which was much higher than the cross-loading of any other language variable. When it was included instead as a dependent variable

Table 2
Canonical Loadings and Weights and Cross-Lagged Correlations

				Cross-lagged correlations								
Task	Load	Coeff	Pretend production	Pretend comprehension	Play narrative	Fantasy	IC	Role play	Language	Frog MS	W3 ToM	W4 ToM
Symbol	.56	.47	.41**	.21	−.08	.17	−.03	.03	.32*	−.15	.26†	.24†
Pretend production	.32	.04	—	.06	.01	.08	.04	.21	.24†	.05	.34*	.16
Pretend comprehension	.24	−.10	—	—	.05	.15	.23†	.21	.18	.18	.21	.23†
Play narrative	.29	.22	—	—	—	.11	−.01	.31*	.20	.13	.04	.22†
Fantasy	.14	−.27	—	—	—	—	.20	.24†	.29*	.16	.30*	.11
IC	.58	.51	—	—	—	—	—	.32*	.09	−.02	.25†	.30*
Role play	.52	.24	—	—	—	—	—	—	.20	.26*	.21	.30*
Language	.69	.44	—	—	—	—	—	—	—	.27*	.31*	.44**
Frog MS	.07	−.01	—	—	—	—	—	—	.12	.17	—	
W3 ToM	.49	.25	—	—	—	—	—	—	—	.31*	—	
W4 ToM	.97	.90	—	—	—	—	—	—	—	—	—	

Note. IC = imaginary companion; Frog MS = mental state terms in child's retelling of frog story; W3 ToM = Wave 3 theory of mind score; W4 ToM = Wave 4 theory of mind score.
†$p < .10$. *$p < .05$. **$p < .01$. (two-tailed)

with ToM, the percentage of variance explained was still significant, $R = .51$ for the first canonical variate, Wilks's $\lambda = .71$, $\chi^2(9) = 16.73$, $p = .05$, and the task had a loading of .69, about equal to the Wave 3 ToM loading of .66. In this case the cross-loading was much lower at .35, similar to the cross-loading of the other ToM variables. These findings suggest that sentential complements has significant ToM and language components, supporting with data a conceptual point others have also made (e.g., Slade & Ruffman, 2005).

Discussion

The cross-lagged, cross-domain investigation of how various symbolic skills might underlie ToM is discussed first, followed by discussion of other findings.

Canonical Modeling of the Symbolic Underpinnings of ToM

Our primary goal in this study was to examine symbolic antecedents to ToM at ages 4 and 5. Among those antecedents were a set of more purely symbolic tasks, tests of language, and measures concerning pretend play. Our canonical model explained 38% of the variance ($R = .62$) between the symbolic antecedents and ToM performance at 4 and 5 years of age, with notable loadings of six symbolic antecedent variables (as shown in Table 2, column 2)—one tapping language across the first three waves, four tapping pretend play across those waves, and one tapping symbolic understanding at the first two waves more directly. The more directly symbolic variable was the sum of two tasks given at 30 and 36 months, respectively: DeLoache's point picture and scale model tasks. Only one other study to our knowledge has shown that performance on just one of these, the scale model task, relates to ToM (Walker & Murachver, 2012). In that study performance on the scale model task at 30 months was not related to ToM at 42 or 48 months, probably due to floor performance on the former. We used point picture at 30 months and had more variability. For our overall model, we combined point picture with scale model because they were strongly interrelated. Our wider array of ToM and symbol tasks showed notable relations at both age 4 and age 5, which is impressive for the predictive power of the symbol tasks because our cross-lagged correlations were over a longer time span.

Four pretend tasks loaded well on the vector that was related to the vector underlying ToM at 4 and 5 years of age: pretend production at 2.5 years, completing a pretend play narrative at 3, role play at 4, and having had an IC. These results dovetail nicely with those of Taylor and Carlson (1997). Pretend play is often thought to contribute to ToM because children take on roles during pretend play, thus practice simulating or seeing the world from different points of view (Harris, 2000; Kavanaugh, 2006; Lillard, 2001). Lillard (2001) proposed that early pretend play with the mother could facilitate both the symbolic function and more pretending, which in turn would—both directly and through pretend role play—contribute to ToM. The interrelations among these tasks strongly support this model, but the canonical analysis emphasizes the symbolic element underlying all the tasks because the underlying vector also derived from relations with language and the more purely symbolic task. Indeed, pretend comprehension and liking fantastical play did not load well on this vector, suggesting an underlying variate that conceptually appears tied to producing symbolic relations was more closely tied to the variate underlying ToM. One possibility concerns generativity of pretend play, often including producing language to describe one's pretend actions or a character's viewpoint. The two role play tasks clearly involve this (talking on the phone, and playing doctor), as does picking up dolls to complete a narrative (where higher scores derived from attributing agency to the dolls, making them act on their own). At an earlier age, pretend production scores also entailed generating pretense.

On the other hand, pretense comprehension and fantasy orientation were not well related to this underlying vector—in fact, their coefficients were negative. Recall that fantasy orientation was the parent's report of what the child most liked to play, alone and with others. Although this was related to language, having an IC, and ToM at 30 months, the reason for these relations appears to be stemming from some other source. Likewise, pretend comprehension was significantly correlated with point picture and vocabulary at Wave 1, and even shows a trend-level correlation with ToM at age 5, but is not related to the underlying variate. One possibility is that pretend comprehension and the symbol tasks share *too* much of the same variance pertaining to the underlying symbolic construct, but symbol is more closely aligned to the underlying construct, overshadowing the contribution of pretend comprehension. This seems possible as both tasks involve

observing an experimenter who sets up a symbolic relation, and then executing some additional step that rests on understanding what one just saw. In contrast, pretend production, role play, the narrative task, and IC involved generating a symbolic relation in various ways, and therefore contributed uniquely to the construct underlying the symbol–ToM relation.

Language significantly loaded on the underlying variate that was strongly related to the construct underlying ToM at Waves 3 and 4. Although many other studies have shown language and ToM relations concurrently and over time (Astington & Baird, 2005; Hughes et al., 2005), our study is unique in highlighting the symbolic interrelations through CC analysis. Walker and Murachver (2012) also examined language using the TELD. They found significant correlations between the TELD at 30, 36, and 42 months and ToM at 48 months, with language mediating the effect of scale model performance on ToM.

Other Within-Wave and Cross-Wave Findings

Wave 1

Within Wave 1, although carried to some degree by age, a notable pattern of relations was revealed between point picture and language production, language comprehension, and pretend comprehension, and (as a trend) pretend production. To our knowledge this is the first study to show such relations to the point picture task. The relation to language echoes Walker and Murachver's (2012) findings with the scale model task at older ages, whereas the relation from the symbol task to pretend play is new. The relation between point picture and pretense comprehension is very interesting; both tasks involve interpreting someone else's intention that one object be taken as a symbol for something else. This supports DeLoache's (2004) interpretation.

The relation within the MB-CDI subscales (production and complexity) replicates prior studies; for example, Fenson et al. (1994) reported correlations of .85 between the two MB–CDI subscales. The lack of relation between language and pretend play production might seem to require explanation, as several studies do show relations between the two skills (see summary in Lillard et al., 2013). However, on close inspection, our result is consistent with prior research. Tamis-LeMonda and Bornstein (1994), for example, also had no correlation between pretend production in free play (using scoring similar to ours) and the Bates Language Inventory at 20 months. Instead, they found a relation between pretend play and the semantic diversity of children's language, which was not measured here. Other studies have seen relations at younger ages (Lyytinen, Laakso, Poikkeus, & Rita, 1999), and have suggested that play and language have different trajectories by age 2 (Dixon & Shore, 1993). Although both variables load well on a common construct related to later ToM, they show no interrelation at 30 months.

The relation between pretend production and pretend comprehension at these ages has not to our knowledge been previously investigated. Leslie (1987) claimed that production of one's own pretense acts and comprehension of others' pretense acts must develop simultaneously, but the complete absence of correlation found here suggests otherwise. The levels of pretend comprehension observed were consistent with past research for children just under 2.5 years of age (Harris & Kavanaugh, 1993; Ma & Lillard, 2006), and although our pretend comprehension and production measures are not directly comparable, our results (showing ample production with comprehension lagging) are consistent with other research showing that pretend production precedes comprehension (Hopkins et al., 2014). Although one must be cautious in drawing interpretations from negative results, the finding does seem problematic for Leslie's (1987) claim of simultaneity.

Wave 2

The most notable finding in Wave 2 was the relation between scale model and PPVT. This again reiterates the findings of Walker and Murachver (2012), but using a different language task than they had used. The TELD obtained a stronger correlation with scale model (.57) than the PPVT obtained (.31), perhaps because the TELD covers more aspects of language and thus a finer appreciation of speaker intent.

Wave 3

The phone and role play tasks were well correlated, suggesting a common ability to act out an imaginary phone call and a doctor–patient (or doctor–nurse) scenario with one's mother; the combined variable representing role play was significantly related to both earlier language skills (MCDI) and later ToM (Wave 4). Also consistent with Taylor and Carlson (1997), having an IC was

significantly related to ToM at age 5, and showed a trend at age 4. Although the CELF had no relation to other tasks, sentential complements was significantly related to ToM. Its relation in fact was so strong that it cross-loaded quite highly on the construct underlying ToM in the main analysis. Further analyses suggest that sentential complements is more a language than a ToM task, yet has elements of both, as others have suggested (Slade & Ruffman, 2005).

Wave 4

The relations between ToM and the TEC were not surprising, as emotion understanding is an aspect of ToM and several items on the TEC presuppose mental state understanding. It is of interest that mental state language in retelling a wordless story book was unrelated to ToM.

Limitations of the Study

This study has several limitations. First, some other construct could be the source of common variance among the tasks. The ability to process symbolic information is of course a key component of human (and artificial) intelligence. Our tasks are believed to tap this important aspect of intelligence, but further work could strengthen our claim by showing that another type of intelligence, perhaps practical intelligence, is unrelated.

Other nonsymbolic antecedents could also be responsible. There are also of course many other possible antecedents to ToM, even embedded within the constructs measured. For example, the association between language and ToM, examined here for symbolic underpinnings, could also turn on conversation. Language permits conversation, which is a natural setting for revealing that people represent the world in different ways (Harris et al., 2005). Similarly, Nelson (2005, p. 32) discussed how "mind exchanges" are enabled by language and lead to ToM. Children who have better language can engage in more conversations, which should in turn facilitate their language development along with ToM. Some evidence for this comes from American Sign Language (ASL)-speaking deaf children of parents who do not speak ASL (Peterson & Siegal, 1995). Conversation in such cases is limited, and the children are also slow to develop a ToM. Another pertinent candidate is executive function. Executive function was tested in the Walker and Murachver (2012) study, and as in many studies (see Carlson, Koenig, & Harms, 2013) was related to language and ToM but did not mediate the relation between them. Some recent studies have found relations between executive function and pretend play (Carlson & White, 2013; Hopkins et al., 2014). This skill should be tested in future studies.

In sum, further research should test how other variables uniquely relate to ToM. Issues of overfitting the data arise when more variables are added; thus, variables must be prudently selected and ideally can be reduced to the pertinent constructs.

Another limitation here is that we only tested a sample of mostly well-educated mothers in western Massachusetts. Participants most typically involved in psychology research are not representative of the world's people, and our sample is not even representative of all Americans: Twenty-eight percent of Americans in general had completed a bachelor's degree or higher in 2009 (Ryan & Siebens, 2012), whereas 76% of our mothers (sampled earlier in the 2000s) had. Education and income tend to rise in concert, and two key constructs measured here, language (Cutting & Dunn, 1999; Hart & Risley, 1995) and pretend play (Smilansky, 1968), have been shown to vary by education and/or income level. ToM, particularly false belief (as opposed to desire, perception, and emotion), does not appear to be influenced by income and education levels (Cutting & Dunn, 1999; Garner, Curenton, & Taylor, 2005; Pears & Moses, 2003; cf. Holmes, Black, & Miller, 1996). However, development in low-income American children typically proceeds along the same trajectories as for higher income children; some developments simply occur later. Our prediction is that the relations we found here would be the same in samples with fewer well-educated mothers, but that we would see them emerge on a later time scale.

Summary

The main purpose of this study was to examine the symbolic antecedents of ToM. Rich interrelations were found showing that a construct underlying pretend play, language, and more purely symbolic tasks shares significant variance with a construct underlying ToM. We take these findings as evidence that ToM development has a symbolic substrate (Lillard, 2001) that emerges early in development and is related to individual differences in ToM at ages 4 and 5. This symbolic substrate, relatively understated in discussions of the correlates and predictors of ToM, is likely also present in two other activities and partly explains their relation to ToM. One of these is pretend play. Given that pre-

tend play could also be related to ToM for other reasons, for example, enacting roles, elucidating its joint symbolic underpinning is important. The other is language.

This study confirms and extends the finding that basic language ability is related to ToM, again highlighting the symbolic aspects of each through the CC analysis. When aligned with other variables tapping an underlying symbolic construct, language using different age-appropriate measures from 2.5 to 4 years was significantly related to the construct underlying mental state understanding at ages 4 and 5. Detailing the extent of these interrelations over time is an important contribution of this research.

In sum, this study shows that there is a symbolic construct, tapped by symbol, pretense, and language tasks, which develops over the preschool years, and is significantly related to performance on ToM measures at ages 4 and 5. This confirms widely held assumptions about one important element of ToM understanding: that understanding mental representation is a symbolic act that relies on a more general symbolic capacity that also undergirds language, pretend play, and understanding symbol–referent relations.

References

Apperly, I. A., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*, 953–970. doi:10.1037/a0016923

Astington, J. W., & Baird, J. (Eds.) (2005). *Why language matters for theory of mind*. New York, NY: Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition and Emotion*, *9*, 151–165. doi:10.1080/02699939508409006

Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education & Development*, *12*, 613–628. doi:10.1207/s15566935eed1204_7

Carlson, S. M., Koenig, M., & Harms, M. B. (2013). Theory of mind. *WIREs Cognitive Science*, 1–12, doi:10.1002/wcs.1232

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & White, R. E. (2013). Executive function, pretend play, and imagination. In M. Taylor (Ed.), *The Oxford handbook of the imagination* (pp. 161–174). New York, NY: Oxford University Press.

Cassidy, K. W., Fineberg, D. S., Brown, K., & Perkins, A. (2005). Theory of mind may be contagious, but you don't catch it from your twin. *Child Development*, *76*, 97–106. doi:0009-3920/2005/7601-0007

Chandler, M., & Helm, D. (1984). Developmental changes in the contributions of shared experience to social role-taking. *International Journal of Behavioral Development*, *7*, 145–156.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*, 853–865. doi:0009-3920/99/7004-0004

de Villiers, J. G. (2007). The interface of language and theory of mind. *Lingua*, *117*, 1858–1878. doi:http://dx.doi.org/10.1016/j.lingua.2006.11.006

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell, & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). East Sussex, UK: Psychology Press.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relation between complex syntax and false-belief-understanding. *Cognitive Development*, *17*, 1037–1060.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*, 1556–1557.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, *62*, 736–752. doi:10.1111/j.1467-8624.1991.tb01566.x

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, *8*, 66–70. doi:10.1016/j.tics.2003.12.004

DeLoache, J. S., Simcock, G., & Marzolf, D. P. (2004). Transfer by very young children in the symbolic retrieval task. *Child Development*, *75*, 1708–1718. doi:10.1111/j.1467-8624.2004.00811.x

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*, 568–570.

Dixon, W. E., & Shore, C. (1993). What shall we play? Play content as a vehicle for parent-infant interactions. *Infant Behavior & Development*, *16*, 121–125.

Dunn, L. M., & Dunn, L. (1997). *Peabody Picture Vocabulary Test–III (PPVT–III)*. Circle Pines, MN: American Guidance Service.

Dunn, J., & Hughes, C. (2001). "I got some swords and you're dead!": Violent fantasy, antisocial behavior, friendship, and moral sensibility in young children. *Child Development*, *72*, 491–505. doi:10.1111/1467-8624.00292

Fenson, L., Dale, P. S., Reznick, J. S., & Bates, E. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(Serial No. 242).

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*, 99–103.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development*, *51*(Serial No. 212).

Flavell, J. H., Green, F. L., & Flavell, E. R. (1990). Developmental changes in young children's knowledge about the mind. *Cognitive Development*, *5*, 1–27. doi:10.1016/0885-2014(90)90010-Q

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*, 483–527. doi:10.1016/0885-2014(95)90024-1

Garner, P. W., Curenton, S. M., & Taylor, K. (2005). Predictors of mental state understanding in preschoolers of varying socioeconomic backgrounds. *International Journal of Behavioral Development*, *29*, 271–281. doi:10.1177/0165025054400005 3

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26–37. doi:10.2307/1130386

Haight, W., & Miller, P. J. (1993). *Pretending at home*. Albany, NY: SUNY Press.

Harris, P. L. (1995). From simulation to folk psychology: The case for development. In M. Davies & T. Stone (Eds.), *Folk psychology*, (Vol. 3, pp. 207–221). Cambridge, UK: Blackwell.

Harris, P. L. (2000). *The work of the imagination*. Oxford, UK: Blackwell.

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, *14*, 69–73. doi:10.1111/j.0963-7214.2005.00337.x

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, *58*(Serial No. 231).

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Holmes, H., Black, C., & Miller, S. (1996). A cross-task comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology*, *63*, 263–285. doi:10.1006/jecp.1996.0050

Hopkins, E. J., Smith, E. D., & Lillard, A. S. (2014). *The development of substitute object pretense: The role of executive function and theory of mind*. Manuscript submitted for publication.

Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology*, *26*, 139–142.

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": Preschoolers' talk about mental states in pretend play. *Cognitive Development*, *12*, 381–403. doi:10.1016/S0885-2014(97)90019-8

Hughes, C., Ensor, R., & Marks, A. (2011). Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends. *Journal of Experimental Child Psychology*, *108*, 96–112.

Hughes, C., Jaffee, S., Happe, F., Taylor, A., Caspi, A., & Moffitt, T. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development*, *76*, 356–370. doi:10.1016/j.jecp.2010.07.012

Kavanaugh, R. D. (2006). Pretend play and theory of mind. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 153–166). New York, NY: Taylor & Francis.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*, 412–426. doi:10.1037/0033-295X.94.4.412

Lewis, C., Freeman, N., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. (1996). Social influences on false belief access: Specific sibling influences or general apprenticeship? *Child Development*, *67*, 2930–2947.

Lillard, A. S. (1993). Pretend play skills and the child's theory of mind. *Child Development*, *64*, 348–371. doi:10.2307/1131255

Lillard, A. S. (2001). Pretend play as Twin Earth. *Developmental Review*, *21*, 1–33. doi:10.1006/drev2001.0532

Lillard, A. S. (2011). Mother-child fantasy play. In A. Pellegrini (Ed.), *Oxford handbook of the development of play* (pp. 284–295). New York, NY: Oxford University Press.

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, *139*, 1–34. doi:10.1037/a0029321

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130–1144. doi:10.1111/1467-8624.00597

Lyytinen, P., Laakso, M. L., Poikkeus, A. M., & Rita, N. (1999). The development and predictive relations of play and language across the second year. *Scandinavian Journal of Psychology*, *40*, 177–186. doi:10.1111/1467-9450.00115

Ma, L., & Lillard, A. S. (2006). Where is the real cheese? Young children's ability to discriminate between real and pretend acts. *Child Development*, *77*, 1762–1777. doi:10.1111/j.1467-8624.2006.00972.x

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press/Penguin.

McCune, L. (1995). A normative study of representational play in the transition to language. *Developmental Psychology*, *31*, 198–206. doi:10.1037/0012-1649.31.2.198

Meins, E., Fernyhough, C., Wainwright, R., Gupta, M. D., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development*, *73*, 1715–1726. doi:10.1111/1467-8624.00501

Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26–49). New York, NY: Oxford University Press.

Pears, K., & Moses, L. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development*, *12*, 1–20.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Leekman, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–137.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development*, *65*, 1228–1238. doi:10.2307/1131316

Peterson, C. C., & Siegal, M. (1995). Deafness, conversation, and theory of mind. *Journal of Child Psychology and Psychiatry*, *36*, 459–474. doi:10.1111/j.1469-7610.1995.tb01303.x

Piaget, J. (1926). *The language and thought of the child*. London, UK: Routledge & Kegan Paul.

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York, NY: Norton.

Pons, F., & Harris, P. L. (2000). *Test of Emotion Comprehension: TEC*. Oxford, UK: Oxford University Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526. doi:10.1017/S0140525X00076512

Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology*, *30*, 332–343. doi:10.1016/j.appdev.2008.12.020

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, *34*, 161–174. doi:10.1037//0012-1649.34.1.161

Ruffman, T., Perner, J., & Parkin, L. (1999). How parenting style affects false belief understanding. *Social Development*, *8*, 395–411. doi:10.1111/1467-9507.00103

Ryan, C. L., & Siebens, J. (2012). *Educational attainment in United States: 2009*: *Current population reports*. Washington, DC: U.S. Census Bureau.

Sachet, A. B., & Mottweiler, C. M. (2013). The distinction between role-play and object substitution in pretend play. In M. Taylor (Ed.), *The Oxford handbook of the imagination* (pp. 175–185). New York, NY: Oxford University Press.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, *23*, 117–141.

Smilansky, S. (1968). *The effects of sociodramatic play on disadvantaged preschool children*. New York, NY: Wiley.

Tamis-LeMonda, C., & Bornstein, M. (1994). Specificity in mother-toddler language-play relations across the second year. *Developmental Psychology*, *30*, 283–292. doi:10.1037/0012-1649.30.2.283

Taylor, M. (1999). *Imaginary companions and the children who create them*. Oxford, UK: Oxford University Press.

Taylor, M. (2005). *Role play coding*. Eugene: University of Oregon.

Taylor, M., & Carlson, S. M. (1997). The relation between individual differences in fantasy and theory of mind. *Child Development*, *68*, 436–455. doi:10.2307/1131670

Walker, R. F., & Murachver, T. (2012). Representation and theory of mind development. *Developmental Psychology*, *48*, 509–520. doi:10.1037/a0025663

Walker-Andrews, A., & Harris, P. L. (1993). Young children's comprehension of pretend causal sequences. *Developmental Psychology*, *29*, 915–921. doi:10.1037//0012-1649.29.5.915

Wang, Y., & Su, Y. (2009). False belief understanding: Children catch it from classmates of different ages. *International Journal of Behavioral Development*, *33*, 331–337. doi:10.1177/0165025409104525

Watson, A. C., Painter, K. M., & Bornstein, M. H. (2001). Longitudinal relations between 2-year-olds' language and 4-year-olds' theory of mind. *Journal of Cognition and Development*, *2*, 449–457. doi:10.1207/S15327647JCD0204_5

Wellman, H. M. (2011). Developing a theory of mind. In U. Goswami (Ed.), *Handbook of cognitive development* (pp. 258–284). Oxford, UK: Blackwell.

Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals: Preschool*. San Antonio, TX: Psychological Corporation.

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. doi:10.1016/0010-0277(83)90004-5

Wolf, D. P., Rygh, J., & Altshuler, J. (1984). Agency and experience: Actions and states in play narratives. In: I. Bretherton (Ed.), *Symbolic play* (pp. 195–218). Cambridge, UK: Academic Press.