Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Experimental Child Psychology 87 (2004) 299–319
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

# Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability

Stephanie M. Carlson,[a,*] Louis J. Moses,[b] and Laura J. Claxton[c]

[a] *Department of Psychology, University of Washington, Seattle, WA 98195, USA*
[b] *Department of Psychology, University of Oregon, Eugene, OR 97403, USA*
[c] *Department of Psychology, University of Massachusetts, Amherst, MA 01002, USA*

Received 27 June 2003; revised 26 January 2004

## Abstract

This research examined the relative contributions of two aspects of executive function—inhibitory control and planning ability—to theory of mind in 49 3- and 4-year-olds. Children were given two standard theory of mind measures (Appearance–Reality and False Belief), three inhibitory control tasks (Bear/Dragon, Whisper, and Gift Delay), three planning tasks (Tower of Hanoi, Truck Loading, and Kitten Delivery), and a receptive vocabulary test (Peabody Picture Vocabulary Test [PPVT-3]). Multiple regression analyses indicated that two inhibition tasks (Bear/Dragon and Whisper) were significantly related to theory of mind after accounting for age, receptive vocabulary, and planning. In contrast, the planning tasks did not share unique variance with theory of mind. These results increase our understanding of the specific nature of executive function–theory of mind relations during early childhood.
Published by Elsevier Inc.

*Keywords:* Executive function; Inhibition; Planning; Theory of mind

## Introduction

The executive functions serve to monitor and control thought and action and include skills such as self-regulation, inhibitory control, planning, attentional

* Corresponding author. Fax: 1-206-685-3157.
*E-mail address:* carlsons@u.washington.edu (S.M. Carlson).

0022-0965/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jecp.2004.01.002

flexibility, error correction and detection, resistance to interference, and working memory (Welsh, Pennington, & Groisser, 1991; Zelazo, Carter, Reznick, & Frye, 1997). Advances in executive functioning are increasingly believed to be linked with theory of mind development during the preschool period (Carlson & Moses, 2001; Frye, Zelazo, & Palfai, 1995; Hughes, 1998a; Moses, 2001; Perner & Lang, 2000; Russell, 1996), although the nature of that linkage remains to be specified.

Recent investigations of typically developing children have shown robust correlations between performance on theory of mind and executive tasks, independent of age and intelligence (e.g., Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Frye et al., 1995; Hala, Hug, & Henderson, 2003; Hughes, 1998a, 1998b; Keenan, Olson, & Marini, 1998; Perner & Lang, 2000; Perner, Lang, & Kloo, 2002). In perhaps the most comprehensive study of this kind, Carlson and Moses (2001) assessed 107 preschoolers on 10 inhibitory control tasks and four theory of mind tasks. The inhibition measures consisted of "conflict" tasks requiring the flexible deployment of dominant and subdominant responses and "delay" tasks requiring children to postpone a dominant response. Theory of mind tasks consisted of appearance–reality, two types of false belief, and deception. The inhibitory control and theory of mind task batteries were highly coherent and strongly correlated with one another ($r = .66$). Importantly, the correlation remained significant after partialling age, receptive vocabulary, sex, and additional controls (including number of siblings, a measure of symbolic play, and nonmental state control tasks). Other investigators have reported very similar results using other executive tasks. In a meta-analysis of many of these studies, Perner and Lang (1999) reported an effect size for the relation of 1.08, which is considered strong in terms of Cohen's (1988) widely used criteria.

Although it has been established that a variety of executive skills are closely bound together with a variety of theory of mind skills, one must determine which executive skills are most strongly associated with theory of mind development. Miyake et al. (2000) stressed the importance of fractionating executive functioning into its various components to make it a more theoretically and clinically useful construct. Furthermore, understanding the specificity of the relation is vital to learning where performance breaks down on theory of mind tasks, which executive skills might be implicated in the emergence of mentalizing ability, and (by extension) which deficits may be present in atypical theory of mind development.

Two executive skills hypothesized to be at the heart of the relation between executive function and theory of mind are inhibitory control (Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998; Hala et al., 2003; Hughes, 1998a, 1998b; Russell, 1996) and working memory (Davis & Pratt, 1996; Gordon & Olson, 1998; Keenan et al., 1998). Successful social cognition requires both the ability to hold in mind multiple perspectives (i.e., working memory) and the ability to suppress irrelevant perspectives (i.e., inhibitory control). Several investigations testing these hypotheses have been reported. For example, Hughes (1998a) gave preschool children a battery of executive function tasks, including inhibitory control, set shifting, and working memory as well as false belief and deception tasks. Individual differences in inhibition and working memory were correlated with each other, and both were significantly related to the theory of mind tasks. However, after the effects of age and both verbal and nonverbal intelligence were held constant, working memory was now unrelated to the theory of mind tasks, whereas inhibitory control remained a significant predictor of at least one of them (deception).

Carlson et al. (2002) reported similar results using a different set of inhibition and working memory measures. Inhibitory control was significantly related to false belief performance after partialling age as well as verbal and nonverbal intellectual ability. The same could not be said for the working memory measures, which were related to theory of mind only in raw correlations. Moreover, in Carlson and colleagues' study, the relation between inhibitory control and theory of mind held up over and above working memory performance in addition to the other controls. There was also evidence for specificity within the inhibition construct (see also Carlson & Moses, 2001). Conflict inhibition tasks, which appear to also have a high working memory load (Diamond, Kirkham, & Amso, 2002), were more strongly related to false belief performance (as well as to working memory) than was a delay task that has relatively low working memory demands (for similar results, see Hala et al., 2003). These findings suggest that a *combination* of inhibition and working memory capacity—constructs that might themselves be integrally related (for perspectives on this issue, see Beveridge, Jarrold, & Pettit, 2002; Engle, 2002; Roberts & Pennington, 1996)—plays a crucial role in children's understanding of false belief and deception.

That said, other executive abilities also might be implicated in normative theory of mind development. Planning ability, in particular, is a complex executive skill that develops on a timetable similar to that of theory of mind (Atance & O'Neill, 2001) and might well be related to it. By 2 years of age, children are beginning to talk about future events, and such talk increases over the preschool years (Hudson, Shapiro, & Sosa, 1995; Nelson, 1989). However, experimental evidence indicates that children's ability to sequence future events, such as waking up, getting dressed, and going to sleep, is not well developed until 4 years of age (Friedman, 1990), the very time at which marked changes in theory of mind performance are also occurring.

Consistent with the dual onset of planning and theory of mind skills, Bischof-Köhler (1998; as cited in Perner & Lang, 2000) found evidence for a relation between them. In that study, 3- and 4-year-olds were administered a planning task in which they had to select items to bring along on a shopping trip. Some items were relevant (e.g., wallet), whereas others were irrelevant yet enticing (e.g., flashlight). The same children also participated in a standard unexpected location false belief task. Bischof-Köhler found that children's planning competence was significantly correlated with false belief performance ($r = .39$). Although these results are intriguing, age and verbal ability were not controlled in that study. In addition, there are clear inhibitory requirements to ignoring the attractive items, so one would want to know whether the false belief–planning relation would hold when individual differences in inhibitory control were held constant. It is also difficult to determine the extent to which the shopping task was a pure planning measure given that it is heavily dependent on script-based knowledge.

Frye (2000) proposed that a more specific aspect of planning ability is implicated in theory of mind development. According to Frye and Zelazo's cognitive complexity and control theory, reasoning with embedded rules is critical for success on many executive and theory of mind tasks (Frye, 1999, 2000; Frye et al., 1995). Frye (2000) suggested that the executive function–theory of mind relation is "most relevant to the executive functions of planning and deliberative action" (p. 156). He argued that conditional reasoning is required for planning appropriate actions as opposed to inhibiting inappropriate ones. According to cognitive complexity and control theory, advances in conditional reasoning (i.e., if–if–then relations) enable children to reflect on the problem at hand and allow them to develop an appropriate plan of responding. This same ability is believed to be required for success on theory of mind problems such as the unexpected location false belief task in which children must analyze what a character's plan of action will be under a specific set of embedded conditions.

Frye (1999) listed several instances of indirect evidence against a simple response inhibition account of age-related changes in performance on executive tasks such as the dimensional change card sort developed by Frye et al. (1995). For example, Jacques, Zelazo, Kirkham, and Semcesen (1999) found that 3-year-olds had trouble detecting the errors of a puppet sorting the cards perseveratively according to a previously correct dimension, even when the children themselves took no part in the sorting game. This kind of evidence is taken to suggest that 3-year-olds' difficulty on executive function and theory of mind tasks is not an inhibitory problem but rather one of initially failing to see that the same action can be put to different purposes within the same situation. Hence, executive function and theory of mind failures could reflect a common difficulty in producing or comprehending intentional embedded action plans.

Consistent with this proposal, Hughes (1998b) found a significant relation between preschool children's performance on the Tower of London (Shallice, 1982), a classic executive planning task, and theory of mind, controlling for age and receptive vocabulary. However, there were also significant relations between theory of mind and inhibition (Luria's Hand Game) and attentional flexibility (i.e., set shifting). Because the primary focus in that study was longitudinal prediction between executive function and theory of mind over a 1-year period, it was not reported whether the planning, inhibition, or set-shifting measures accounted for unique variance in theory of mind when entered simultaneously in a multiple regression. The longitudinal analyses showed that one task in particular, detour reaching (an inhibitory control task), best predicted later theory of mind. However, no planning measure was administered at the first time point to help tease apart the role of inhibition and planning ability.

Preliminary evidence suggesting that planning might not in fact be a crucial source of executive function–theory of mind relations comes from Carlson and Moses (2001). They administered a motor sequencing task along with their inhibitory control and theory of mind batteries. This task, adapted from Welsh et al. (1991), requires children to tap the keys of a musical keyboard in sequence over and over again as fast as they can for 10 s. It calls for motor planning but has relatively weak inhibition demands. Motor sequencing was related to both the inhibitory control and theory of mind task batteries, but the correlations fell below significance after age, sex, and receptive vocabulary were controlled. Furthermore, the relation between inhibitory control and theory of mind remained highly significant when motor sequencing was partialled along with the control variables. This finding suggests that executive function–theory of mind relations may be specifically mediated by inhibitory control. However, because only one type of planning task was included, it could not be ruled out that other aspects of executive planning ability would subsume the relation between inhibitory control and theory of mind. In particular, following Frye's (2000) proposal, planning measures that include hierarchically embedded goal-oriented actions, as opposed to simpler action plans such as motor sequencing, ought to provide the most rigorous test of the relation between planning and theory of mind.

In sum, there is evidence suggesting a specific relation between executive function and theory of mind that might be underpinned by inhibitory processes (or inhibitory processes in combination with working memory). However, it remains possible that the relation is not unique to inhibitory control but instead is a product of other executive abilities such as hierarchical planning. Therefore, the goal of this study was to examine the relative contributions to preschoolers' theory of mind of inhibitory control and planning ability. The major question under investigation was whether these executive abilities indeed relate to theory of mind over and above relevant controls and, if so, whether their associations with theory of mind are unique or overlapping.

To answer that question, we administered theory of mind, inhibitory control, and planning batteries, as well as a measure of receptive vocabulary, to a sample of typically developing preschoolers. The theory of mind battery consisted of standard false belief and appearance–reality tasks (Flavell, Flavell, & Green, 1983; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983). The inhibitory measures, all originally derived from Kochanska, Murray, Jacques, Koenig, and Vandegeest (1996), consisted of Bear/Dragon, requiring children to alternately perform and suppress actions requested by puppets; Whisper, requiring them to whisper the names of familiar cartoon characters; and Gift Delay, requiring them not to peek while an experimenter noisily wrapped a gift. We selected these tasks because they had previously been found to correlate with theory of mind and because they represented the two separate aspects of inhibition (conflict and delay) described by Carlson and Moses (2001; see also Carlson et al., 2002).

The planning battery consisted of three established and developmentally appropriate measures from previous research: Tower of Hanoi (Simon, 1975; Welsh, 1991), in which children were required to transfer disks onto pegs according to a set of rules; Truck Loading (Fagot & Gauvain, 1997), in which children had to load party invitations onto a truck in reverse order so as to deliver them in an efficient manner; and Kitten Delivery (Fabricius, 1988), a route planning task in which children were asked to deliver kittens to their mother as quickly as possible. In all three tasks, children needed to ignore local suboptimal solutions and instead plan the series of actions that would allow them to solve the problems most efficiently. Moreover, all three tasks involve conditional reasoning along the lines specified by Frye (2000). For example, in the Truck Loading task, if one is to deliver invitations from the top of the stack, and if the pink house is *last*, the solution is to load the pink invitation onto the truck *first*.

If the executive function–theory of mind relation is specific to inhibitory control, inhibition should relate to theory of mind over and above planning as well as age and receptive vocabulary. The reverse pattern would be expected if planning ability is central to the executive function–theory of mind relation. Finally, if planning and inhibition make independent contributions to theory of mind, both should relate significantly to theory of mind over the controls.

## Method

### Participants

The participants were 49 preschool children from the Seattle, Washington, metropolitan area (mean age = 4 years 0 months, range = 38–59 months). There were 24 3-year-olds (10 boys and 14 girls, mean age = 3 years 7 months) and 25 4-year-olds (12 boys and 13 girls, mean age = 4 years 4 months). One additional child participated but did not complete the study.

### Procedure

Children were tested individually in a single 45-min videotaped laboratory session. As is standard practice in individual differences research, measures were presented in a fixed order (for a rationale, see Carlson & Moses, 2001). The order of tasks was as follows: Peabody Picture Vocabulary Test (third edition, PPVT-3), Appearance–Reality, Tower of Hanoi, Bear/Dragon, Contents False Belief, Truck Loading, Whisper, Location False Belief, Gift Delay, and Kitten Delivery. Each measure is described in detail in the following section. The same female experimenter tested all children.

### Measures

#### Verbal ability measure

Children were given the PPVT-3 (Dunn & Dunn, 1997). The PPVT is a measure of receptive vocabulary that correlates highly with full-scale verbal intelligence measures such as the Wechsler Preschool and Primary Scales of Intelligence (revised, WPPSI-R) (Carvajal, Parks, Logan, & Page, 1992) and the verbal subscale of the Stanford–Binet IV (Hodapp, 1993), as well as with theory of mind (Carlson & Moses, 2001). The experimenter states a word, and children select the picture that best illustrates it (out of four choices). Testing continues until children err on 8 out of a set of 12 words. One 4-year-old did not complete the PPVT-3.

#### Theory of mind measures

*Location False Belief*. In a version of Wimmer and Perner's (1983) standard unexpected location false belief task, two puppets (Bert and Ernie) played with a ball briefly, and then Bert put the ball in a blue container and left. Ernie retrieved the ball, played briefly with it, and then put it away in a red container and left. Finally, Bert returned, wanting to play with the ball, and the experimenter asked the false belief question ("Where does Bert think the ball is?") followed by the reality question ("Where is the ball really?"). Two 3-year-olds and a 4-year-old erred on the reality control question. In this and the other theory of mind tasks, such cases were treated as missing data.[1]

*Contents False Belief*. Following procedures developed by Perner et al. (1987) and Gopnik and Astington (1988), the experimenter presented a Band-Aid box and asked children what they thought was inside. After it was revealed that the box actually contained crayons, the experimenter closed the lid and asked children about their own former false belief ("When you first saw this box, before we opened it, what did you think was inside?"), the belief of a naive puppet ("Here comes Ernie. He has never looked inside this box before. What does he think is inside?"), and the reality control question ("What's really inside?"). Children were scored for their knowledge of their own former belief and the other's current false belief (range = 0–2). Six 3-year-olds erred on the reality control question. Contents False Belief scores were also missing for two 4-year-olds due to experimenter error.

*Appearance–Reality*. The experimenter showed children two objects with misleading appearances (Flavell et al., 1983; Flavell, Green, & Flavell, 1986). One was a piece of sponge painted to look like a rock, and the other was a picture of a red castle that looked black when held behind a green filter. For each stimulus, children first were shown how the object looked and the true identity or true color of the object. Next, the experimenter asked the appearance question ("When you look at this right now, does it look [like a sponge/red] or does it look [like a rock/black]?"), followed by the reality question ("What [is/color is it] really?"). Children received credit on each task if they answered both the appearance and reality questions correctly (range = 0–2). Appearance–Reality data were missing for one 3-year-old due to camera failure.

*Inhibitory control measures*

*Bear/Dragon*. The Bear/Dragon task (Kochanska et al., 1996; Reed, Pien, & Rothbart, 1984) is a simplified version of "Simon Says" in which children need to selectively suppress commanded actions. To begin, the experimenter asked children to imitate 10 self-directed actions (e.g., "Touch your ears"). She then introduced two puppets—a "nice bear" and a "naughty dragon"—and instructed children to do what the bear asked them to do but not to follow the dragon's commands. In practice trials, the experimenter moved the bear's mouth and said (in a high-pitched voice), "Touch your nose," and then moved the dragon's mouth and said (in a low gruff voice), "Touch your tummy." Children passed the practice if they followed the bear's command but ignored the dragon's command. All children but four (three 3-year-olds and one 4-year-old) succeeded on the bear practice the first time. For the

[1] No child answered more than one reality control question incorrectly. The results were similar when children who failed the control question were instead scored as failing the task.

dragon practice, children who failed five practice trials in a row were told that the experimenter would help them on a final practice trial by holding their hands down on the table. Two children (one 3-year-old and one 4-year-old) required this level of assistance on the dragon practice trials. The experimenter then did a verbal rule check by asking children what they should do when the bear asked them to do something, and she repeated the question for the dragon. The experimenter provided feedback on children's responses and asked both questions again as necessary. Five children (four of them 3-year-olds) required correction on the rule checks. (All passed on the second trial.) This was followed by 10 test trials (5 bear trials and 5 dragon trials in alternating order) in which children were given no assistance. They were reminded of the rules after 5 trials regardless of performance. Children received scores ranging from 0 to 3 on each dragon trial (0 = *a full commanded movement*, 1 = *a partial commanded movement*, 2 = *a wrong movement*, 3 = *no movement*). Four 3-year-olds refused to complete the task. Reliability coding was conducted on a randomly selected 33% of the sample ($n = 16$). Disagreements were resolved by a third coder. For the rule check, Cohen's kappa = 1.0; for test trials, kappa = .80.

*Whisper*. This task required children to whisper during an exciting identification game (Kochanska et al., 1996). To warm up, the experimenter asked children to whisper their names. Most participants were able to do so on the first try. Then, they were asked to whisper the names of 10 cartoon characters consecutively presented on laminated cards. Of these characters, 6 were familiar to most preschoolers at the time the study was conducted (e.g., Big Bird, Mickey Mouse), and 4 were unfamiliar to children of this age (e.g., Fat Albert). The experimenter spoke in a whisper throughout and reminded children to whisper after the first five trials. Scoring was as follows: 0 = *a shout*, 1 = *a normal or mixed voice*, and 2 = *a whisper.* Interrater reliability was high for practice and test trials, Cohen's kappas = 1.0 and .97, respectively.

*Gift Delay*. This measure called for delay of gratification (Kochanska et al., 1996). The experimenter asked children to sit in a chair facing away and to try not to look while she wrapped a gift for them. She then noisily wrapped the gift over a period of 60 s. Finally, she invited children to open their present (a small toy animal). Coding included (a) a peeking score (0 = *turning fully around to peek*, 1 = *peeking over the shoulder*, 2 = *no attempt to peek*), (b) the total number of times children peeked, and (c) latency to peek over the shoulder. Coder agreement was as follows: peeking score, Cohen's kappa = .80; total number of peeks (exact agreement), kappa = .63 (all were within a score of 1); and latency to peek was within 1 s on 94% of double-coded cases.

*Planning measures*

*Tower of Hanoi*. Children were given Welsh's (1991) simplified version of this classic planning task developed by Simon (1975). The simplified version is more appropriate for younger children. It included an ascending order of difficulty across trials. In addition, it allowed children to execute a move sequence without first verbalizing their plans, and children received explicit feedback on their errors. Following Welsh (1991), this was labeled the "Monkey Jumping Game" in which three wooden disks represented monkeys (smallest disk = baby sister monkey, medium-sized disk = boy monkey, and largest disk = daddy monkey), three wooden pegs represented trees, and the table represented a huge river around the trees. The experimenter explained that the larger monkeys could not sit on top of the smaller monkeys in the same tree because they would "smush" them, but the smaller monkeys could sit on top of the larger monkeys. Children were told that only one monkey could jump at a time and that the monkeys could never be put on the table because that would mean that they fell into the water. In a rule check that followed, most children demonstrated that they understood the rules on the first try, with seven children (five of them 3-year-olds) requiring one extra reminder before passing the rule check. Next, the experimenter introduced a second set of trees and monkeys (the experimenter's monkeys) and told children that the children's monkeys were copycat monkeys that always wanted to look like her monkeys. The task increased in difficulty, starting with two wooden disks requiring two moves to match the experimenter's monkeys (which were always positioned on the far right goal peg) and eventually progressing to using three wooden disks requiring four moves to match the experimenter's monkeys. Altogether, there were six levels of difficulty. For each level of difficulty, children received two trials and had to pass one of the two trials to continue to the next level. The experimenter reminded children of the relevant rule after each incorrect trial (e.g., that only one monkey can jump at a time). Data were not recorded for one 3-year-old due to camera failure, and one 4-year-old did not want to play this game. Performance was scored as the highest level of planning successfully completed (0 to 6), Cohen's kappa = 1.0.

*Truck Loading*. In this adaptation of a task developed by Fagot and Gauvain (1997), children were asked to pretend that they were mail carriers using a toy mail truck to deliver differently colored party invitations to similarly colored wooden houses placed around a street block (a large poster board depicting a neighborhood). The experimenter first demonstrated the game using one house and one invitation. She instructed children to load the invitation onto the back of the truck and then explained that the neighborhood had a one-way street as she demonstrated the route to be taken by the truck. Directional arrows were marked on the road as a reminder. Children then completed a warmup with two houses in which they were given the following instructions: "Now there are two houses that we want to invite to the party. The yellow invitation goes to the yellow house, and the purple invitation goes to the purple house. Now, we need to deliver these party invitations *fast* so that everyone will be able to come to the party. The fastest way is to drive around the block only one time." Next, the experimenter suggested a way in which to put the invitations on the back of the truck so that they could be delivered quickly. "You always have to take the letter off the *top* of the truck so that the top invitation goes to the first house and the next invitation goes to the next house." She demonstrated loading the truck in reverse order. Then, children were given a rule check ("So, can I take one from the *bottom* of the truck?"). The experimenter repeated this until children answered correctly ("no"), giving feedback after each answer. All but 12 children needed correction on the order rule. After children delivered these invitations with the experimenter's help, the warmup houses were replaced with two differently colored houses (red and black) and the test trials began. One new house was added for each successive level of difficulty, ending with five houses and a total of four levels of difficulty. For each level, children received two trials and had to pass one of them to continue to the next level. The dependent variable at each level was whether or not children stacked the invitations in reverse order to that in which they needed to be delivered. Self-corrections were permitted during the loading phase. After loading, children proceeded to deliver the invitations and the experimenter gave feedback about the relevant rule on incorrect trials (e.g., that children must always take the invitation from the top of the truck). Then, she collected the invitations and returned the truck to the starting point to try again (up to two trials at each level). Children received scores based on the highest level achieved (0 to 4). Coding was reliable, Cohen's kappa = 1.0.

*Kitten Delivery*. This task was derived from one used by Fabricius (1988). Children are required to plan to minimize the distance traveled in gathering kittens located in buckets around the room. Wellman, Fabricius, and Sophian (1985) first developed this method to help separate nonplanful "sighting" from planful search techniques. These earlier studies showed that planning was clearly evident in children's search sequences by $3\frac{1}{2}$ years of age and improved over the remaining preschool years. In a warmup, two buckets containing one toy kitten each were placed together in the center of the room, and a toy mother cat was placed 12 in behind them. The experimenter explained that the kittens had been playing in the buckets, had gotten stuck, and now needed help in getting out of the buckets because the mother cat was really worried about her kittens. Then, the experimenter asked children to remove the kittens from the buckets and to set them next to the mother. There were two levels of difficulty in the main task (two kittens and three kittens). For each level, children received two trials and had to pass one of them to continue to the next level. For Level 1, the kittens were set in buckets on opposite sides of the room, 90 in apart. Children were positioned at the far end of the room so that they were 90 in away from each bucket; thus, the children and buckets formed an equilateral triangle. For Level 2, the third bucket was positioned 113 in across from the children and 60 in away from each of the other two buckets; thus, the three buckets formed an isosceles triangle. The position of the mother cat alternated between the far left bucket and the far right bucket on each trial. Children were told that the kittens were playing in the buckets again and had gotten stuck. On each trial, the refrain from the experimenter was, "You'll want to stop by each bucket and pick up each kitten and bring it to the mama cat. You want to go the *quick way* so you don't have to do a lot of walking and you can get the baby kittens to the mother cat right away. She needs them as soon as possible because she's really worried. Okay, go ahead and bring the kittens to their mother as soon as you can." For Level 1, the most efficient route was to go to the bucket farthest from the mother first and then to the bucket beside the mother. For Level 2, the most efficient route was to go to the bucket farthest from the mother first, then to the middle bucket, and last to the bucket beside the mother. In contrast to the preceding planning tasks, the experimenter did not provide explicit feedback about inefficient solutions but simply reminded children of the objective to go quickly and not to do a lot of walking at the beginning of each trial. Children were scored according to the highest level achieved (0, 1, or 2). Coding agreement was high, Cohen's kappa = .90.

## Results

We first describe the results for the individual measures, followed by the major analyses of the relations among them.

### *Vocabulary assessment*

The average age-standardized score on the PPVT-3 was 111.8, $SD = 15.0$. Descriptive statistics for the raw scores are displayed in Table 1. As shown in Table 2, raw PPVT-3 scores were correlated with age but not with sex. Raw scores were used in subsequent analyses that included age as a covariate.

### *Theory of mind assessment*

We first computed average Appearance–Reality scores (across the sponge/rock and castle trials) and average False Belief scores (across location, contents [self], and contents [other] measures) for each participant. Children's performance on the theory of mind measures in each age group is shown in Table 1. As indicated, 4-year-olds performed better than 3-year-olds, significantly so on Appearance–Reality. As shown in Table 2, Appearance–Reality and False Belief were significantly related. Therefore, we computed composite scores (average across the five theory of mind items) for use in further analyses. Scores were prorated for missing data. The theory of mind composite was significantly correlated with age and PPVT-3 but was unrelated to sex, and the same was true for the False Belief and Appearance–Reality measures taken separately (Table 2).

### *Executive function assessment*

#### *Inhibitory control*

Mean scores on each of the inhibitory measures are shown in Table 1. Although 4-year-olds performed better than 3-year-olds on all three tasks, the difference was significant only for the Whisper test trials. Following Carlson and Moses (2001), for each inhibitory control task, we created a single score by standardizing and aggregating across the dependent measures that were significantly intercorrelated: on Bear/Dragon, the number of dragon practice trials (reversed) and dragon test trial scores, $r(45) = .85$; on Whisper, the number of practice trials (reversed) and mean scores, $r(49) = .54$; and on Gift Delay, all three dependent measures of waiting, $rs(49) = .73$ to $.86$ (all $ps < .001$). As shown in Table 2, two of the aggregated inhibitory control measures—Bear/Dragon and Gift Delay—were significantly correlated

Table 1
Performance on all measures as a function of age

| Measure | 3-year-olds ($n = 24$) | 4-year-olds ($n = 25$) | Total ($N = 49$) | Age effects |
|---|---|---|---|---|
| Receptive vocabulary | | | | |
| PPVT-3 | 55.46 (16.23) | 73.29 (20.47) | 64.38 (20.37) | $t(46) = 3.34^{**}$ |
| Theory of mind | | | | |
| Appearance–Reality (range = 0–1) | .41 (.36) | .66 (.37) | .54 (.38) | $t(46) = 2.33^{*}$ |
| False Belief (range = 0–1) | .35 (.41) | .49 (.36) | .42 (.39) | $t(47) = 1.33$ |
| Composite (range = 0–1) | .36 (.29) | .57 (.32) | .46 (.32) | $t(47) = 2.37^{*}$ |
| Inhibitory control | | | | |
| Bear/Dragon | | | | |
| Practice required (range = 1–6) | 1.72 (1.24) | 1.40 (1.08) | 1.50 (1.15) | $t(44) = 0.64$ |
| Dragon score (range = 0–15) | 12.40 (4.65) | 13.72 (3.21) | 13.13 (3.92) | $t(43) = 1.12$ |
| Whisper | | | | |
| Practice required (range = 1–2) | 1.13 (0.34) | 1.08 (0.28) | 1.10 (0.31) | $t(47) = 0.51$ |
| Mean score (range = 1–2) | 1.87 (0.29) | 1.99 (0.04) | 1.93 (0.21) | $t(47) = 2.05^{*}$ |
| Gift Delay | | | | |
| Peek score (range = 0–2) | 1.08 (0.93) | 1.28 (0.74) | 1.18 (0.83) | $t(47) = 0.82$ |
| Total peeks (range = 0–6) | 1.84 (1.95) | 1.36 (1.44) | 1.59 (1.71) | $t(47) = 0.97$ |
| Latency (range = 1–60 s) | 35.75 (25.40) | 45.42 (18.90) | 40.68 (22.62) | $t(47) = 1.51$ |
| Planning | | | | |
| Tower of Hanoi (range = 0–6) | 2.00 (1.68) | 2.17 (1.97) | 2.09 (1.82) | $t(45) = 0.31$ |
| Truck Loading (range = 0–4) | 1.38 (1.01) | 2.24 (1.51) | 1.82 (1.35) | $t(47) = 2.35^{*}$ |
| Kitten Delivery (range = 0–2) | 1.17 (0.70) | 1.12 (0.83) | 1.14 (0.76) | $t(47) = 0.21$ |

*Note.* Standard deviations are in parentheses. The *N*s for individual tasks ranged from 41 to 49 due to missing data.

$^{*}p < .05$.

$^{**}p < .01$.

with one another. Whisper scores were unrelated to these two tasks. As a result, we analyzed the inhibitory control tasks separately in further analyses. Bear/Dragon scores were significantly correlated with age (Table 2). No other relations between inhibitory control and age, sex, and PPVT-3 were significant.

*Planning*

Mean scores on the planning measures are shown in the bottom portion of Table 1. As indicated, 4-year-olds outperformed 3-year-olds significantly on the Truck Loading task. Intercorrelations among the planning measures are provided in Table 2. Tower of Hanoi and Truck Loading were significantly related. Kitten Delivery was unrelated to these two tasks. Following our approach with respect to the inhibitory control measures, we examined the three planning tasks separately in subsequent analyses. As shown in Table 2, Truck Loading scores were

Table 2
Correlations between measures

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *(a) Bivariate correlations* | | | | | | | | | | | |
| 1. Age | −.02 | .50** | .36* | .24 | .20 | .15 | .57** | −.08 | .42** | .40** | .50** |
| 2. Sex | | −.22 | .02 | .16 | .24 | −.10 | .00 | .01 | .05 | −.02 | .00 |
| 3. PPVT-3 | | | .25 | .26 | .18 | .32* | .42** | .10 | .54** | .34* | .49** |
| 4. Bear/Dragon | | | | −.05 | .36* | .04 | .36* | .01 | .32* | .43** | .41** |
| 5. Whisper | | | | | .08 | .23 | .19 | .23 | .34* | .31* | .37** |
| 6. Gift Delay | | | | | | .18 | .19 | .21 | .25 | .21 | .22 |
| 7. Tower of Hanoi | | | | | | | .41** | .23 | .37* | −.06 | .15 |
| 8. Truck Loading | | | | | | | | .03 | .48** | .34* | .48** |
| 9. Kitten Delivery | | | | | | | | | .12 | .03 | .06 |
| 10. Appearance–Reality | | | | | | | | | | .41** | .79** |
| 11. False Belief | | | | | | | | | | | .88** |
| 12. Theory of Mind Composite | | | | | | | | | | | |

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| *(b) Partial correlations controlling for age and receptive vocabulary* | | | | | | | | |
| 1. Bear/Dragon | −.17 | .31* | −.04 | .19 | .03 | .18 | .32* | .32* |
| 2. Whisper | | .02 | .17 | .04 | .24 | .37* | .22 | .34* |
| 3. Gift Delay | | | .13 | .08 | .22 | .16 | .13 | .05 |
| 4. Tower of Hanoi | | | | .36* | .21 | .25 | −.18 | .02 |
| 5. Truck Loading | | | | | .06 | .28 | .13 | .19 |
| 6. Kitten Delivery | | | | | | .12 | .04 | .07 |
| 7. Appearance–Reality | | | | | | | .29† | .70** |
| 8. False Belief | | | | | | | | .84** |
| 9. Theory of Mind Composite | | | | | | | | |

*Note.* *N*s ranged from 42 to 47.
* $p < .05$.
** $p < .01$.
† $p = .05$.

significantly correlated with age and PPVT-3. Tower of Hanoi performance was also related to PPVT-3. Kitten Delivery was not related to age or PPVT-3. There were no significant correlations with sex.

Prior to further correlational analyses, we examined whether there were floor effects on any of these planning tasks. This was not the case. The proportions of children who produced perfect solutions at the simplest level of difficulty were 81, 92, and 78% for Tower of Hanoi, Truck Loading, and Kitten Delivery, respectively. This analysis suggests that children's failures at the more complex levels are likely due to their increased planning demands rather than to mere confusion or low-level difficulties in understanding the tasks.

*Relation between inhibitory control and planning*

Next, we examined whether our executive function measures themselves were related and whether this relation would remain after we controlled for age and receptive vocabulary. If so, this would raise the possibility of a general executive function–theory of mind correlation but not a specific contribution by either inhibition or planning. As shown in Table 2, only one such relation reached statistical significance, that is, that between Bear/Dragon and Truck Loading. After partialling age and PPVT-3, however, this relation fell below significance. Therefore, planning and inhibitory control appeared to be largely independent constructs as measured by the tasks included here.

*Specifying the relation between executive function and theory of mind*

The next and most critical series of analyses was aimed at specifying the relative contributions of inhibitory control, planning, age, and vocabulary to theory of mind. As shown in Table 2, the correlations between theory of mind and two of the inhibitory control measures (Bear/Dragon and Whisper) were significant. Importantly, these relations remained significant after controlling for effects due to age and receptive vocabulary. The same general pattern was apparent for False Belief and Appearance–Reality taken separately, although some of the partial correlations fell below significance.

The correlations between the planning measures and theory of mind are also shown in Table 2. Truck Loading was significantly related to the Theory of Mind Composite as well as to the False Belief and Appearance–Reality measures in the raw correlations, and Tower of Hanoi was related to the False Belief measure. However, unlike the findings for inhibitory control, these relations did not hold up over and above age and PPVT-3.

*Regression analyses*

To determine the specific contribution of inhibitory control to theory of mind when planning ability was accounted for in addition to the controls, we carried out a series of hierarchical multiple regressions. In the first of these, with the Theory of Mind Composite as the dependent variable, we included the control variables (age and PPVT-3) in the first block. Age was a significant predictor and PPVT-3 was marginally significant, $\beta s = .43$ and .27, $ts(42) = 3.02$ and 1.91, $ps < .01$ and .07, respectively. The three inhibitory control measures were entered simultaneously in the second block. Consistent with the partial correlation analyses we reported, both Bear/Dragon and Whisper were significant predictors of theory of mind performance, $\beta s = .32$ and .31, $ts(5, 37) = 2.41$ and 2.63, $ps < .025$ and .015, respectively. In the third block, we entered the three planning measures simultaneously. The results of the final model are shown in Table 3. Even with all three planning measures included, the Bear/Dragon and Whisper tasks remained significant predictors of theory of mind. In contrast, the planning measures failed to account for significant variance in theory of mind scores over and above the inhibition measures and controls.

In the second regression analysis, we reversed the order of entry such that the planning measures were entered in the second block (after age and PPVT-3) and

Table 3
Hierarchical multiple regression of variables predicting theory of mind (final models)

| Variable | $\beta$ | $t$ | $p$ |
|---|---|---|---|
| *(a) Theory of Mind Composite as the criterion*[a] | | | |
| Age | .15 | 0.87 | .39 |
| PPVT-3 | .26 | 1.91 | .07[†] |
| Bear/Dragon | .31 | 2.28 | .03[*] |
| Whisper | .30 | 2.36 | .02[*] |
| Gift Delay | .03 | 0.20 | .85 |
| Tower of Hanoi | −.07 | −0.52 | .61 |
| Truck Loading | .17 | 1.04 | .31 |
| Kitten Delivery | .00 | 0.03 | .98 |
| *(b) Appearance–Reality as the criterion*[b] | | | |
| Age | .00 | −0.03 | .98 |
| PPVT-3 | .34 | 2.55 | .02[*] |
| Bear/Dragon | .24 | 1.73 | .09[†] |
| Whisper | .30 | 2.40 | .02[*] |
| Gift Delay | .07 | 0.56 | .58 |
| Tower of Hanoi | .16 | 1.18 | .25 |
| Truck Loading | .18 | 1.13 | .27 |
| Kitten Delivery | .00 | −0.01 | .99 |
| *(c) False Belief as the criterion*[c] | | | |
| Age | .22 | 1.10 | .28 |
| PPVT-3 | .15 | 0.95 | .35 |
| Bear/Dragon | .28 | 1.79 | .08[†] |
| Whisper | .23 | 1.58 | .13 |
| Gift Delay | .00 | 0.01 | .99 |
| Tower of Hanoi | −.21 | −1.39 | .17 |
| Truck Loading | .08 | 0.44 | .67 |
| Kitten Delivery | .01 | 0.07 | .94 |

[a] *Note.* $N = 42$. Multiple $R = .73$; $R^2 = .54$; adjusted $R^2 = .43$; $SE = .24$.
[b] *Note.* $N = 42$. Multiple $R = .74$; $R^2 = .54$; adjusted $R^2 = .43$; $SE = .30$.
[c] *Note.* $N = 42$. Multiple $R = .62$; $R^2 = .38$; adjusted $R^2 = .24$; $SE = .33$.
[*] $p < .05$.
[†] $p = .05$.

inhibitory control was entered last. None of the planning measures was significantly related to theory of mind in the second block: Tower of Hanoi, Truck Loading, and Kitten Delivery ($\beta$s $= -.06$, .21, and .06, respectively). Thus, there was no evidence that planning was a significant predictor of theory of mind, even without controlling for inhibition.

Next, we addressed the question of whether a similar pattern of results would hold when using each individual theory of mind measure in turn as the criterion variable. It remained possible that planning ability would play a stronger role than would inhibition with respect to the Appearance–Reality and False Belief measures. These results are shown in Table 3. In a regression analysis predicting Appearance–Reality from age, PPVT-3, the three inhibitory control measures, and the three planning measures, both PPVT-3 and Whisper scores were significant in the final model. Bear/Dragon was marginally significant. Age, Gift Delay, and the three planning measures were not significant. In a parallel analysis using False Belief as the criterion, in the final model, only Bear/Dragon was a marginally significant predictor. In both of these analyses, the planning measures all were nonsignificant even when entered second, that is, after age and PPVT-3 but before the inhibition measures. Thus, these results were broadly consistent with the findings for overall theory of mind: individual differences in inhibitory control, but not planning, were related to theory of mind performance.

Finally, in an effort to allow planning–theory of mind relations to emerge if they existed, we made two changes to our main regression analysis. First, the Kitten Delivery planning task not only failed to relate to theory of mind but also did not relate to the other planning tasks or to age and vocabulary. Although the task has been used successfully as a planning measure in previous research (Fabricius, 1988; Wellman et al., 1985), it would be reasonable to question its validity as an index of planning in the current study. Given that, the measure may have simply been muddying the waters in our regression analysis. Second, in contrast to the Kitten Delivery measure, the other two planning measures were significantly related to one another. Therefore, we reconducted our main regression analysis without the Kitten Delivery measure and with a new composite planning measure that aggregated Truck Loading and Tower of Hanoi. However, even in this analysis, the same pattern emerged. The Whisper and Bear/Dragon measures were again significant predictors of theory of mind performance over age and PPVT-3 ($p$s $< .05$), whereas the new composite planning measure did not approach significance ($p > .65$).

## Discussion

The remarkable changes in children's theory of mind during the preschool period do not occur in a vacuum. Instead, they coincide with a number of other important advances in cognitive development during that period, most notably (from our perspective) executive functioning. Several aspects of executive function have been linked with theory of mind, including inhibitory control and working memory (Carlson & Moses, 2001; Carlson et al., 2002; Davis & Pratt, 1996; Gordon & Olson, 1998; Hala et al., 2003; Hughes, 1998a, 1998b; Keenan et al., 1998). The goal of this investigation was to begin to further specify which components of executive functioning (a rather heterogeneous set of cognitive problem-solving skills) are most strongly associated with theory of mind development. We focused, in particular, on inhibitory control and planning ability. Both of these skills undergo dramatic improvement during the preschool period, and both have been implicated in theory of mind development.

Our findings indicated that two inhibition tasks, in particular, were significantly related to children's theory of mind performance over and above effects due to age and receptive vocabulary: Bear/Dragon and Whisper. A similar pattern of results was obtained for the Appearance–Reality and False Belief tasks taken separately. These results are consistent with other reports suggesting a close association between individual differences in inhibitory control and theory of mind using these and other inhibition tasks (e.g., Carlson & Moses, 2001; Carlson et al., 2002; Frye et al., 1995; Hala et al., 2003; Hughes, 1998a, 1998b; Perner & Lang, 2000). In studies that included the same inhibitory control measures used in the current study, Carlson and Moses (2001; see also Carlson et al., 2002) reported that Bear/Dragon and Whisper were more strongly related to theory of mind than was Gift Delay. In Carlson and Moses (2001), these measures also loaded onto two different factors (conflict and delay) in a principal components analysis. Conflict tasks such as Bear/Dragon and Whisper require children to suppress a dominant response that conflicts with the response called for in the experimental context. In the case of Bear/Dragon, children must initiate and suppress a prepotent response in an alternating pattern. On the Whisper task, they must initiate a subdominant response across a series of varying trials throughout the game. On Gift Delay and similar tasks, in contrast, children need to postpone a prepotent response and remain "in idle" for a period of time. Carlson and Moses (2001) proposed that what may separate the conflict and delay tasks is working memory load: conflict tasks require that children hold in mind the pertinent rules as well as inhibit a prepotent response. Consistent with this proposal, Carlson et al. (2002) reported that Bear/Dragon and Whisper (conflict tasks) were significantly related to working memory tasks and to False Belief, whereas Gift Delay was not. The differential relations to theory of mind found for the inhibitory tasks included in the current study are consistent with Carlson and Moses' analysis.

In contrast, none of the planning measures was significantly related to theory of mind or its subcomponents (False Belief and Appearance–Reality) in parallel analyses controlling for age and receptive vocabulary. Intriguingly, this was the case even though the Truck Loading planning task had exhibited the strongest raw correlation with theory of mind of any of the executive measures. However, Truck Loading also showed the strongest relations to age and vocabulary of these measures. Given that the theory of mind tasks were also strongly correlated with age and vocabulary, it appears that the relation between Truck Loading and theory of mind was driven largely by general cognitive/maturational factors as opposed to some more specific factor common to these constructs. Furthermore, regression analyses revealed that the inhibition measures contributed to theory of mind significantly even when the three planning measures were controlled in addition to age and PPVT-3. These findings are consistent with those of Carlson and Moses (2001), who reported that a motor sequencing task was unrelated to theory of mind after partialling age, sex, and verbal ability. The current study significantly extends those results by examining a larger and more complex set of planning measures. It also extends Hughes's (1998b) research by demonstrating that the conflict inhibition measures were related to theory of mind even when individual differences in planning were controlled.

We selected planning tasks that had the common feature of requiring children to develop and execute a plan involving one or more embedded actions. Following Frye (2000), we hypothesized that this aspect of planning would be the strongest candidate to test in relation to theory of mind. A further advantage of using measures of this type is that such measures, unlike Bischof-Köhler's shopping task, do not rely on children's script-based knowledge. Nonetheless, as noted previously, we failed to find a relation between planning and theory of mind. It is important to note that the results were not due to floor effects on the planning measures. Even the 3-year-olds in our study exhibited an ability to find the first-level solution to the hierarchical planning problems. Moreover, our participants demonstrated the full range of scores on each of the planning measures, suggesting that restricted variability cannot explain the lack of correlations with theory of mind. It was also not the case that the effects of planning were masked by shared variance with inhibitory control given that the planning and inhibition tasks were not significantly related when age and PPVT-3 were controlled. In addition, it seems unlikely that shared information processing and/or linguistic demands, as opposed to inhibitory factors per se, could account for the differential pattern of relations between our executive tasks and theory of mind. For example, it is not obvious that the planning tasks imposed fewer memory demands than did the inhibition and theory of mind measures. Furthermore, although it is true that the planning tasks all called for nonverbal responses, it is also the case that one of the significant predictors of theory of mind (Bear/Dragon) shared this characteristic.

That said, we included only three measures of planning, and of course it is possible that other planning tasks might show a different pattern of relations to theory of mind. Moreover, planning is a multidimensional construct, and we focused on only a specific subset of planning tasks, albeit those that have been argued to be most likely to relate to theory of mind. With that in mind, future research might include a larger and more varied battery of planning measures (e.g., measures that more strongly emphasize future-oriented thinking and/or verbalized plans). Nonetheless, our results suggest that sequential embedded action planning might not be the key factor underlying advances in both executive functioning and theory of mind.

In conclusion, a growing body of research indicates that individual differences in inhibitory control and theory of mind are closely bound together during development. Earlier research has found that this relation persists over age, sex, intelligence, sibling status, and symbolic play. In addition, the relation is especially strong for conflict inhibition measures as opposed to delay inhibition measures, and it remains when other aspects of executive function, such as working memory capacity and motor sequencing, are controlled (Carlson & Moses, 2001; Carlson et al., 2002; Hala et al., 2003; Hughes, 1998a, 1998b; Perner et al., 2002). The findings of the current research add to this evidence by showing that the relation persists over executive planning ability as well. Together, these various findings strongly point to conflict inhibition as being at the heart of the executive function–theory of mind relation.

## Acknowledgments

We thank Serena Word and Antoinette Wong for assistance with data collection and coding.

## References

Atance, C., & O'Neill, D. K. (2001). Planning in 3-year olds: A reflection of future self. In C. Moore & K. Lemon (Eds.), *The self in time: Developmental perspectives* (pp. 121–140). Mahwah, NJ: Lawrence Erlbaum.

Beveridge, M., Jarrold, C., & Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant and Child Development, 11*, 107–123.

Bischof-Köhler, D. (1998). *Projekt zur Untersuchung der spezifischen menschlichen verhaltensorganisation: Zeitreprasentation, theory of mind, und motivationsmanagement bei vierjahrigen*. Unpublished manuscript, University of Munich.

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

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Carvajal, H. H., Parks, J. P., Logan, R. A., & Page, G. L. (1992). Comparisons of the IQ and vocabulary scores on Wechsler Preschool and Primary Scale of Intelligence—Revised and Peabody Picture Vocabulary Test-Revised. *Psychology in the Schools, 29*, 22–24.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Davis, H. L., & Pratt, C. (1996). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25–31.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary task* (3rd ed.). Circle Pines, MN: American Guidance Service.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Fabricius, W. V. (1988). The development of forward search planning in preschoolers. *Child Development, 59*, 1473–1488.

Fagot, B. I., & Gauvain, M. (1997). Mother–child problem solving: Continuity through the early childhood years. *Developmental Psychology, 33*, 480–488.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance–reality distinction. *Cognitive Psychology, 15*, 95–120.

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* (1, Serial No. 212).

Friedman, W. J. (1990). Children's representations of the pattern of daily activities. *Child Development, 61*, 1399–1412.

Frye, D. (1999). Development of intention: The relation of executive function to theory of mind. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self control* (pp. 119–132). Mahwah, NJ: Lawrence Erlbaum.

Frye, D. (2000). Theory of mind, domain specificity, and reasoning. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 149–167). Hove, UK: Psychology Press.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

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.

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*, 70–83.

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development, 4*, 275–298.

Hodapp, A. F. (1993). Correlation between Stanford–Binet IV and PPVT-R scores for young children. *Psychological Reports, 73*, 1152–1154.

Hudson, J. A., Shapiro, L. R., & Sosa, B. B. (1995). Planning in the real world: Preschool children's scripts and plans for familiar events. *Child Development, 66*, 984–998.

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

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*, 1326–1339.

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology, 35*, 770–780.

Keenan, T., Olson, D. R., & Marini, Z. (1998). Working memory and children's developing understanding of mind. *Australian Journal of Psychology, 50*, 76–82.

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development, 72*, 688–690.

Nelson, K. (1989). *Narratives from the crib*. Cambridge, MA: Harvard University Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*, 337–344.

Perner, J., & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship?. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (pp. 150–181). Oxford, UK: Oxford University Press.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73*, 752–767.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty understanding false belief: Representational limitation, lack of knowledge, or pragmatic misunderstanding? *British Journal of Developmental Psychology, 5*, 125–137.

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill–Palmer Quarterly, 30*, 131–147.

Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology, 12*, 105–126.

Russell, J. (1996). *Agency: Its role in mental development*. Mahwah, NJ: Lawrence Erlbaum.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Part B, 298*, 199–209.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 268–288.

Wellman, H. M., Fabricius, W. V., & Sophian, C. (1985). The early development of planning. In H. M. Wellman (Ed.), *Children's searching: The development of search skill and spatial representation* (pp. 123–149). Hillsdale, NJ: Lawrence Erlbaum.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development, 6*, 59–76.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.

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.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*, 1–29.