Developmental Psychology
2013, Vol. 49, No. 9, 1615–1627

© 2012 American Psychological Association
0012-1649/13/$12.00 DOI: 10.1037/a0031056

Individual Differences in Executive Functioning Predict Preschoolers' Improvement From Theory-of-Mind Training

Jeannette E. Benson and Mark A. Sabbagh
Queen's University

Stephanie M. Carlson and Philip David Zelazo
University of Minnesota

Twenty-four 3.5-year-old children who initially showed poor performance on false-belief tasks participated in a training protocol designed to promote performance on these tasks. Our aim was to determine whether the extent to which children benefited from training was predicted by their performance on a battery of executive functioning tasks. Findings indicated that individual differences in executive functioning performance strongly and consistently predicted improvement in children's false-belief performance and their ability to appropriately explain false-belief-based behavior, both during the training period and during the posttest. These findings were robust after statistically controlling for several relevant covariates. These results are consistent with the suggestion that executive functioning skills promote developments in theory of mind by facilitating the ability to reflect upon and learn from relevant experience.

Keywords: theory of mind, false belief, executive functioning, training study, cognitive development

Skillful negotiation of the social world relies in large part on having a "theory of mind"—the capacity to understand that others' actions are motivated by internal mental states (Wellman, 1990). Within the literature, considerable emphasis has been placed on studying advances in children's understanding that beliefs can be false. Although recent research has suggested that some implicit understanding of false belief may be present in very young children (e.g., Onishi & Baillargeon, 2005), an explicit representational theory of mind (RTM) seems to show a more protracted timeline, with major developments occurring between the ages of 3 and 5 years. The precise timeline on which an RTM emerges is related to both experiential factors (e.g., parent–child talk about mental states) and domain-general cognitive developments (e.g., executive functioning). The goal of this study was to examine one way in which experiential and domain-general cognitive factors might interact in shaping RTM development. Specifically, we examined the association between individual differences in executive functioning and children's benefit from training designed to improve false-belief understanding.

Executive functioning refers to the processes that underlie goal-directed behavior, including self-regulation, planning, working memory, response inhibition, and resistance to interference (Carlson, Zelazo, & Faja, in press). The association between executive functioning and RTM has been well established, with the bulk of research showing that children's facility with response conflict executive functioning (RC-EF) tasks involving a dominant response conflict (e.g., Stroop-like tasks) is a necessary prerequisite for reasoning about false belief (see Benson & Sabbagh, 2009, for a recent review). Less well understood, however, is what role RC-EF plays in RTM task performance. One proposed explanation is that RTM tasks such as the false-belief task have inherent RC-EF demands, and that children's maturing RC-EF helps them to negotiate these demands (see e.g., Carlson, Moses, & Hix, 1998; Leslie & Polizzi, 1998). One implication of this view—sometimes called the "expression" account—is that failure on RTM tasks may not reflect a lack of RTM understanding, but rather a lack of sufficient RC-EF.

Here, we investigate an alternate view—sometimes called the "emergence" account–that RC-EF plays a critical role in allowing children to capitalize on the experiences that are relevant to gaining competence with RTM reasoning in everyday contexts. This view is consistent with the role that RC-EF appears to play in promoting learning across a number of domains, including mathematics and language (e.g., Blair & Razza, 2007; Bull & Scerif, 2001; Espy et al., 2004). For instance, Espy et al. (2004) found that preschoolers' inhibitory control skills were predictive of mathematical abilities, even after controlling for age, vocabulary, ma-

---

Jeannette E. Benson and Mark A. Sabbagh, Department of Psychology, Queen's University, Kingston, Ontario, Canada; Stephanie M. Carlson and Philip David Zelazo, Institute of Child Development, University of Minnesota.

This article is based on data from Jeannette E. Benson's master's thesis at Queen's University. This research was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant awarded to Mark A. Sabbagh and by a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship awarded to Jeannette E. Benson. We thank Valerie Kuhlmeier, Tom Hollenstein, and Elizabeth Kelley for helpful feedback, both during the planning stages of this project and during article write-up. We also thank those who assisted with data collection: Katie Gamble, Ashley Hobb, Grace Liao, Anna Lom, and Linda van der Meer. Finally, we would like to thank the families from the Kingston, Ontario, community who graciously volunteered their participation.

Correspondence concerning this article should be addressed to Jeannette E. Benson, Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada. E-mail: j.benson@queensu.ca or sabbagh@queensu.ca

ternal education, and other aspects of executive control (e.g., working memory and shifting abilities). Our suggestion here is that just as RC-EF facilitates learning in other domains, RC-EF may play the same type of role in promoting RTM competence during the preschool years.

Empirical Reasons to Favor an “Emergence” Account

Two bodies of evidence provide support for the “emergence” account described above. The first is that some key predictions of the expression account are contradicted in the literature. For one, although false-belief tasks certainly do have nontrivial RC-EF demands, a meta-analysis by Wellman, Cross, and Watson (2001) showed that manipulations that lower the RC-EF demands of false-belief tasks do not raise children’s performance to above-chance levels. Moreover, RC-EF abilities correlate with performance on these false-belief tasks with lowered RC-EF demands (Henning, Spinath, & Aschersleben, 2011). Also, the neuromaturational changes that are associated with RTM development during the preschool years occur within regions that are typically associated with RTM reasoning (e.g., dorsal-medial prefrontal cortex, right temporal parietal juncture) and not regions that are associated with RC-EF (e.g., cingulate cortex, rostro-lateral prefrontal cortex, Sabbagh, Bowman, Evraire, & Ito, 2009). Finally, there are several cultural groups (e.g., Chinese and Korean) that show an accelerated timetable of RC-EF development relative to North Americans, yet no parallel advantage in RTM reasoning (Oh & Lewis, 2008; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Thus, it appears that preschoolers’ RTM performance (or “expression”) is not entirely determined by achieving threshold levels of executive competence.

The second body of evidence favoring the “emergence” account includes longitudinal findings showing that RC-EF appears to contribute to developments in RTM reasoning. In a microgenetic study, RC-EF was a developmental precursor to false belief (Flynn, O’Malley, & Wood, 2004), and longitudinal studies thus far show that RC-EF skills predict subsequent variance in RTM better than the reverse developmental ordering (and independent of child general cognitive ability and socioeconomic factors). This is the case in normative (Carlson, Mandell, & Williams, 2004; Hughes, 1998a), low-income (Hughes & Ensor, 2005), and autistic samples (Pellicano, 2007). On the emergence view, this evidence suggests that RC-EF plays a role in the development of false-belief understanding; if RC-EF skills were necessary for false-belief expression (rather than emergence), then we would not expect to find evidence for a unidirectional, longitudinal relationship between RC-EF and subsequent false-belief knowledge.

The “Theory” Theory, RC-EF, and the Role of Experience in RTM Development

Insofar as preschoolers’ RTM development reflects a genuine conceptual change in understanding of minds, the “theory” theory is a commonly invoked framework for both characterizing these changes and proposing mechanisms for how changes might occur (Gopnik & Wellman, 1994). According to the theory theory, mental state understanding originates with nonrepresentational understandings of mental states such as desires and intentions (Gopnik & Wellman, 1994). Over time, children come to recognize that these nonrepresentational theories occasionally lead to incorrect predictions, and in response to these errors, they revise their original theories to minimize error. What the emergence account adds to the theory theory is an account of the role of domain-general neurocognitive skills in the process of theory change. In particular, the emergence account suggests that the relation between RC-EF and false belief exists in part because RC-EF skills predict children’s ability to change their naïve understandings in accordance with false-belief-relevant experiences.

It is well established that RTM development is associated with experiential factors thought to influence children’s exposure to false-belief relevant information. For instance, RTM development is associated with parent–child talk about mental states (e.g., Ruffman, Slade, & Crowe, 2002) and a range of demographic factors, including having older siblings in the home (e.g., Ruffman, Perner, Naito, Parkin, & Clements, 1998), maternal education (Cutting & Dunn, 1999), and socioeconomic status (Hughes & Dunn, 1998; Pears & Moses, 2003). The positive effects of experience on RTM development have also been shown in more controlled laboratory studies that use training regimens to highlight and explain preschoolers’ mistakes in false-belief tasks (e.g., Amsterlaw & Wellman, 2006; Hale & Tager-Flusberg, 2003).

Our proposal is that children’s RC-EF skills may play a critical role in facilitating the process of learning from experiences in order to develop an RTM. There are at least two general mechanisms by which this facilitation might occur. First, children with advanced RC-EF skills may be better able to engage in naturalistic social interactions—both at home and in structured learning settings—that provide them with more exposure to relevant experiences from which to learn (e.g., Hughes, 1998b). Second, RC-EF skills might contribute to learning by facilitating the absorption of information once children are experiencing both formal and informal interactions. Indeed, RC-EF is associated with various cognitive processes that likely play a critical role in learning from feedback, including, (a) identifying and attending to relevant variables (e.g., Diamond, Barnett, Thomas, & Munro, 2007; Garon, Bryson, & Smith, 2008), (b) noticing discrepancies between expectations and subsequent outcomes (i.e., error monitoring, Zelazo, Carlson, & Kesek, 2008), (c) increasing the probability of reflection on response choices (Carlson et al., in press; Zelazo, 2004), and more speculatively, (d) flexibly updating prior knowledge based upon new information.

The Present Study

The goal of the present study was to test the hypothesis that individual differences in preschoolers’ RC-EF skills contribute to the process of developing an RTM from relevant feedback. Although there are empirical and theoretical reasons to suspect this to be true, no studies to date have directly tested this hypothesis. In designing this study, we had to make certain methodological choices. First, given the obvious difficulties associated with measuring the effects of RC-EF on the broad and abstract nature of RTM, we opted to examine only children’s understanding that people will search for objects in the last place that they were seen and will thus search unsuccessfully when the object has been unexpectedly moved. Although an RTM certainly encompasses more than this, prior research has shown that an understanding of this particular entailment can be trained in a short period when children are given appropriate feedback in the false-belief context, but not otherwise (Hale & Tager-Flusberg, 2003). Thus, in our study children with varying RC-EF skills were exposed to a 2-week training regimen designed to teach them that people search for objects where they last left them. We hypothesized that individual differences in RC-EF would predict training-related advances in children's reasoning about this particular facet of false-belief understanding. A more open question was whether the training would affect children's reasoning about other aspects of false-belief understanding (e.g., that many different types of experiences can cause false beliefs, which in turn can influence a wide variety of behaviors).

Method

Participants

Participants were 32 preschool-aged children (16 boys, 16 girls). At the time of initial testing, children ranged in age from 3 years, 6 months, to 3 years, 11 months (M = 3 years, 9 months, SD = 1.57 months). Three additional children were recruited but excluded from analyses because they did not complete one or more of the study sessions ($n = 2$) or because of distraction in the home ($n = 1$). Eight of the 32 participants (4 boys, 4 girls) were included in preliminary analyses to establish the reliability of the measures but not focal analyses because they showed strong performance (50% correct or higher) on the false-belief battery prior to training. The final sample therefore consisted of 24 children (12 boys, 12 girls, M age = 3 years, 8 months, SD = 1.61 months). Participants were recruited through local-area daycares and through a database comprised of families recruited through community events. All children were native English speakers drawn from a primarily White, middle-class community.

Design and Procedure

Children were tested individually across four videotaped sessions, including an initial testing session, two training sessions (each involving four false-belief training tasks), and a final testing session. Each session took approximately 20–30 min, and the same female experimenter administered all study sessions. The four sessions were conducted within a 2- to 3-week period, with a minimum of 2 days elapsing between each session. Based on each family's preference, the study sessions were held in the children's daycare ($n = 4$), in the laboratory ($n = 1$), or in a quiet room in the family's home ($n = 19$). During the initial testing session, children were presented with a false-belief battery, two RC-EF tasks, and three measures of pre-false-belief theory-of-mind understanding. The second session consisted of the first four false-belief training tasks and the Peabody Picture Vocabulary Test–Third Edition (PPVT-III). During the third session, the remaining four false-belief training tasks were administered along with three additional RC-EF measures. During the final testing session, the false-belief battery and RC-EF tasks from the initial testing session were readministered. The tasks were administered in a fixed order for all children (see Table 1). Using a fixed order ensured that individual differences in children's performance could not be attributed to variance in task administration. All of the tasks administered here have been used in other studies, so only a brief description of each is included below.

Materials and Measures

False-belief battery. Children received this battery at both the initial and final testing sessions. For each task, surface characteristics (i.e., the characters and objects involved) were changed across testing sessions.

Contents change (Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987). Children were shown a container that normally would contain something familiar (e.g., a crayon box) and asked to state what they thought it contained. The experimenter then revealed that the box actually contained something unexpected (e.g., candles). Children were then introduced to a naïve character and asked what she would think the box contained. This question was followed by a control question: "Did [character's name] see inside the box?" Children were given a score of 1 if they answered both the test and memory questions correctly and a score of 0 otherwise.

Location change (Wimmer & Perner, 1983). A character was shown to hide an object in one location and leave the scene. In his or her absence, a second character took the object from its original location and hid it elsewhere. The first character then returned, and children were asked a control question: "Did [character's name] *see* the [object] get moved?" If the children responded incorrectly ($n = 12$), the story was reenacted and the question was asked

Table 1
Task Administration Sequence

Session 1	Session 2	Session 3	Session 4
Contents change	False Belief Training 1	False Belief Training 5	Contents change
Location change	False Belief Training 2	DCCS	Location change
Appearance/reality	False Belief Training 3	False Belief Training 6	Appearance/reality
Deceptive pointing	False Belief Training 4	Whisper	Deceptive pointing
Bear/Dragon	PPVT-III	False Belief Training 7	Bear/Dragon
Grass/Snow		KRISP	Grass/Snow
Diverse desires		False Belief Training 8	
Diverse beliefs			
Knowledge access			
Real–apparent emotion			

Note. PPVT-III = Peabody Picture Vocabulary Test–Third Edition; DCCS = Dimensional Change Card Sort; KRISP = Kansas Reflection–Impulsivity Scale for Preschoolers.

again. Finally, the test question was asked: “When [character’s name] comes back, where will he look for the [object]?” (score: 0–1).

***Appearance-reality* (*Flavell, Green, & Flavell, 1986*).** Children were shown two objects that had misleading appearances. The first was an object that appeared to be one thing, but was actually another (e.g., a sponge that looked like a rock). The second was a picture that appeared to be one color when covered by a transparent screen but was actually another (e.g., an orange castle that looked black under a screen). Children were asked first how the object looks “when you look at it right now,” and second, what the true nature of the object was. On each trial, children received a score of 1 if they correctly responded to both test questions (total score: 0–2).

***Deceptive pointing* (*Carlson et al., 1998*).** Children were told that a puppet was going to look for an object and that it would be fun to play a trick by pointing to make him look in an empty box. Children participated in two trials that each involved playing a trick on a different character. On each trial, children were awarded a score of 1 if they successfully used deception by pointing to the empty box (total score: 0–2).

RC-EF battery. The Dimensional Change Card Sort (DCCS) task was administered once during the second training session. The Bear/Dragon and Grass/Snow tasks were administered at both the initial and final testing sessions. We administered these tasks twice for two reasons. The first was that performance on these two tasks is usually strongly related to false-belief understanding, and because of our small sample size, having two administrations of each offered the opportunity for a stable measure of these skills. The second was to assess and control for the extent that RC-EF skills may have improved over the training period. Although our hypothesis was that children’s RC-EF *prior* to training affects their ability to learn from the training, an alternative explanation is that training has salutary effects on RC-EF, which in turn affect false-belief performance. By administering these two tasks twice, we were able to assess this possibility.

Two additional RC-EF tasks—the Kansas Reflection–Impulsivity Scale for Preschoolers and Whisper tasks (see Carlson & Moses, 2001)—were administered during the third training session but excluded from analyses because children as a group showed near ceiling performance on the tasks.

***Bear/Dragon* (*Kochanska, Murray, Jacques, Koenig, 1996; Reed, Pein, & Rothbart, 1984*).** This task is a modified version of a “Simon Says” game. Children were asked to follow the directions given to them by a “nice bear” puppet and not to follow the directions of a “mean dragon” puppet. Directions involved clearly defined actions (e.g., touch your head). The puppets were both controlled by the experimenter, and bear trials alternated with dragon trials. Prior to testing, children participated in practice trials during which they were given feedback until they responded correctly to one bear trial and one dragon trial. Responses on the dragon test trials were coded on a scale of 0–4, where 0 represented committing a full commanded movement, and 4 represented no movement in response to the command. Interrater reliability was established by having a second independent coder who was blind to the study’s hypotheses code 30% of the sample. Interrater reliability was high for both the practice trials (agreement = 95.24%) and the test trials (agreement = 97.89%, Cohen’s κ = .96).

***Grass/Snow* (*Carlson & Moses, 2001*).** A black board with a white card attached to the upper left-hand corner and a green card attached to the upper right-hand corner was placed in front of the children. The experimenter instructed them to play a silly game by pointing to the green square anytime she said the word “snow” and to the white square anytime she said the word “grass.” Practice trials where children were given feedback on their answers were administered until children responded correctly to one grass trial and one snow trial consecutively. Following this, 16 test trials were administered where no feedback was provided. On each of the test trials, children were given a score of 1 for initially pointing to the correct square and 0 for either initially pointing to the incorrect square or pointing to both squares at the same time. Interrater reliability (established as above) was high for the practice trials (agreement = 94.74%) and the test trials (agreement = 96.11%, Cohen’s κ = .91).

***Dimensional Change Card Sort* (*Frye, Zelazo, & Palfai, 1995; Zelazo, 2006*).** Two boxes were placed in front of the children, one showing a picture of a red rabbit and the second a picture of a blue boat. Children were sequentially handed cards that varied on two dimensions: color and shape (i.e., red and blue rabbits and boats). For the first five cards, children were asked to sort based on shape (i.e., rabbit cards go in the box with the rabbit on the front, boat cards go in the box with the boat on the front). For the last five cards, children were asked to switch rules and sort based on color (i.e., red cards go in the box with the red shape on the front, blue cards go in the box with the blue shape on the front). Two of the postswitch cards were compatible with the initial rule and three were incompatible (i.e., following the shape rule would lead to incorrect responses). Children were given a score of 1 for each correct postswitch trial (total score: 0–5).

False-belief training. The training procedure was modeled after a false-belief training paradigm developed by Hale and Tager-Flusberg (2003) and shown to be effective with young preschool-aged children. In each training session, the experimenter enacted four location change stories using toy figurines and props. At the end of each story, children were asked to predict where a character with a false belief would search for an object. Incorrect responses were given corrective feedback comprised of a simple explanation and story reenactment. Correct responses were confirmed, and a shortened explanation was provided with no reenactment. Characters and surface features of the stories were changed across all false-belief tasks administered. Further, the direction of object transfer (from the left side of the scene to the right, or vice versa), the characters’ positions, and the direction of the characters’ departures from the scene were counterbalanced across training trials to prevent children from learning a simple rule to pass the tasks based on the location or movement of characters or objects. A sample false-belief training script is presented in the Appendix.

Explanation coding. Many researchers have highlighted the importance of testing children’s false-belief knowledge using both prediction and explanation measures (e.g., Bartsch, 1998). During the second training session, children were asked to provide explanations for their answers on each of the four training trials following their response to the training task test question. Children’s explanations were classified into the following categories: (a) False-belief Relevant: reference to a story fact that constitutes a

Table 2
Types of False Belief Explanations

Explanation	Examples
False-belief relevant	"That's where it was when he left." "Cause that's where he left it when he leaved."
False-belief irrelevant	"It's in there." (pointing to object's current location) "He won't find his seashell." "Cause the man moved [the object] in [the object's current location]."
Uninterpretable	"Because his shell." "Because he should." "Because, he'll look in there and there and there (pointing to multiple locations)."
Don't know (or no response)	"I don't know."

reasonable explanation for the character's actions, (b) False-belief Irrelevant: reference to a story fact that is irrelevant to the character's actions, (c) Uninterpretable, and (d) Don't Know/No Response (see Table 2 for examples). Only explanations following the 50 instances where children provided a correct test question response were coded. Interrater reliability was 94.00% (Cohen's κ = .91), with the two raters coding 47 of the 50 explanations in the same way. Disagreements were resolved through discussion.

Distraction coding. The ability to control attention might be one mechanism by which RC-EF affects children's ability to learn from feedback. To explore this possibility, children's behavior during administration of the false-belief training tasks was coded to assess attention to the presented information. We limited our coding to portions of the false-belief training tasks that were the same across all children in the focal sample. Distraction was thus coded from the start of each training task until the point at which the test question was posed. Measures of distraction included (a) irrelevant speech (making an off-topic comment), (b) touching or grabbing the stimuli, and (c) looking away for 3 or more consecutive seconds. Interrater reliability calculated on 30% of the coded segments was high (irrelevant speech: agreement = 95.3%, Cohen's κ = .859; touching the stimuli: agreement = 95.3%, Cohen's κ = .685; looking away: agreement = 100%, Cohen's κ = 1.0).

Pre-false-belief theory-of-mind measure (Wellman & Liu, 2004). Tasks included in the measure assessed children's understanding of a set of mental state understandings that reliably emerge prior to false-belief knowledge. Tasks were administered in a fixed order of ascending difficulty. Each task required that children reason about a character's mental state that was highlighted to be different from their own. The diverse desires task required children to predict what snack a character—who was said to have different preferences than the child—would choose. The diverse beliefs task required children to predict where a character would look for an object when the object's true whereabouts were unknown and the character was said to have a different belief about the object's whereabouts than the child. The knowledge access task tested the understanding that a character who has never seen inside a nondescript box before will not know its contents. Toy figurines, puppets, and picture props were used to illustrate the task scenarios. For each task, correct responses were awarded a score of 1. We included this measure to control for the possibility that children with stronger RC-EF also showed more advanced pre-false-belief theory-of-mind understanding and were thus better prepared to benefit from training (scores: 0–3).

Language measure. The Peabody Picture Vocabulary Test–Third Edition (PPVT-III, Dunn & Dunn, 1997) is a standardized measure of children's receptive language abilities. This measure was included to control for general language skills when examining individual differences in children's performance across tasks. In each trial, children were shown a page with four line drawings and asked to point to the line drawing that corresponded with a vocabulary word spoken by the experimenter. Children continued through progressively more difficult items until they incorrectly answered more than eight in a set of 12, or until they completed all sets of words in the task. Raw scores on the measure were calculated by subtracting the total number of errors made from the highest item answered.

Results

Our primary goal was to determine whether individual differences in children's RC-EF skills prior to training predicted the extent to which children's false-belief performance benefited from training. We begin by reporting results from preliminary analyses on our measures independently. We then report the focal analyses examining the extent to which our measures of false-belief improvement were related to RC-EF. Only participants who demonstrated poor performance on the initial false-belief battery participated in Sessions 2–4. Thus, all analyses (with two exceptions, discussed below) were run with this focal sample of 24 participants.

RC-EF

The RC-EF battery was comprised of three tasks: Bear/Dragon, Grass/Snow, and DCCS. Descriptive statistics for these tasks are presented below and are highly consistent with Carlson (2005), who reported on a large sample of children in this age range.

Bear/Dragon. Both administrations of Bear/Dragon (before and after training) involved two dependent measures: number of tries taken to pass the practice trials and percentage score on the test trials (see Table 3). There was a significant decline in the

Table 3
Performance on Tasks Administered at Both Initial and Final Testing Sessions

Task	Initial testing period *M* (*SD*)	Final testing period *M* (*SD*)
False belief battery		
Location change (out of 1)	.08 (.28)	.75 (.44)
Contents change (out of 1)	.09 (.29)	.13 (.34)
Appearance/reality (out of 2)	0.58 (0.65)	0.67 (0.76)
Deceptive pointing (out of 2)	0.75 (0.79)	1.17 (0.87)
Bear/Dragon		
Practice trials (number of trials taken)	2.79 (2.65)	2.00 (2.82)
Test trials (% score)	66.45 (37.49)	70.85 (36.02)
Grass/Snow		
Practice trials (number of trials taken)	3.59 (2.26)	2.55 (1.10)
Test trials (% score)	52.44 (31.24)	61.94 (23.07)

number of practice trials needed across sessions, $t(23) = 2.26$, $p = .034$, $d = 0.31$, but the difference in performance on test trials was not significant, $t(23) = 1.00$, *ns.* The lack of a difference on the test trials suggests that children's Bear/Dragon performance was stable over time. Indeed, aggregate initial and aggregate final Bear/Dragon scores were strongly related, $r(22) = .83$, $p < .001$. An aggregate Bear/Dragon score was calculated by combining standardized total scores on the dragon practice (total number of practice trials until correct responding, reverse scored) and test trials from both testing sessions, following Carlson and Moses (2001).

Grass/Snow. Both administrations of Grass/Snow involved two dependent measures: number of tries taken to pass the practice trials and percentage score on the test trials (see Table 3). As with Bear/Dragon, paired *t* tests showed that the decline in practice trials across time was significant, $t(21) = 2.71$, $p = .013$, $d = 0.65$, but the difference in test trial performance was not, $t(22) = 1.22$, *ns.* The correlation between aggregate initial and final Grass/Snow scores was moderate, but just below standard significance levels, $r(22) = .39$, $p = .069$. An aggregate Grass/Snow score was calculated by combining standardized total scores on the practice (number of trials until correct responses, reverse scored) and test trials from both testing sessions. These findings in combination with the Bear/Dragon findings strongly suggest that RC-EF was stable across time and that participation in the training had little effect on children's RC-EF performance.

DCCS. Children were scored on five postswitch trials where they were asked to sort cards according to color after previously having to sort according to shape. All children responded correctly on the two compatible postswitch trials where following either the new rule or the old rule resulted in the same sorting outcome. Percentage scores were relatively low on the remaining three incompatible postswitch trials where children had to switch flexibly between dimensions in order to successfully sort by the new rule (color; $M = 27.67\%$, $SD = 44.67\%$).

RC-EF Scale aggregating. Intercorrelations of the standardized Bear/Dragon, Grass/Snow and DCCS tasks were moderate and in line with previous studies (rs $= .360$ to $.426$, $\alpha = .65$). Thus, standardized scores on the three measures were aggregated to form a single measure of children's overall RC-EF skills.

False-Belief Improvement Measures

False-belief battery. The false-belief battery—including the Contents Change, Location Change, Deceptive Pointing, and Appearance/Reality tasks—was administered during both the initial and the final testing sessions. Scores on these tasks are presented in Table 3. Proportions on each task (0 to 1.00) were used instead of sums when forming initial and final false-belief battery scores to give equal weight to each of the four tasks. Because poor performance on this measure was used as a screening tool to be included in the focal sample, reliability analyses performed on the artificially restricted performance of this group would be uninformative. For the full sample of 32 children, the individual tasks were moderately correlated with one another (rs $= .417$ to $.597$, $\alpha = .74$), giving confidence that performance on the false-belief battery was in line with prior research.

Results of a paired samples *t* test on the focal sample comparing initial and final false-belief battery scores showed that children performed significantly better on the second administration of the battery in comparison to their initial performance, $t(23) = 4.20$, $p < .001$, $d = 1.15$. This demonstrates that children's false-belief performance improved over the training period. However, closer inspection of the results showed that children only showed improvement on two of the four tasks: Location Change, which was the focus of training, $t(23) = 6.78$, $p < .001$, $d = 1.80$, and Deceptive Pointing, $t(23) = 2.46$, $p = .020$, $d = 0.50$.

False-belief training. A repeated-measures analysis of variance (ANOVA) showed that overall, children's performance steadily improved across the training tasks, $F(7, 161) = 5.61$, $p < .001$; linear contrast, $F(1, 23) = 27.90$, $p < .001$. Improvement was surprisingly rapid; none of the children responded correctly on the first training trial, whereas 33.33% did so on the second. By the final training trial, 58.33% of the children responded correctly to the test question.

Control Variables

No significant differences between girls' and boys' scores were found on any of the individual dependent measures (ts $= 1.30$ to 1.85, all ps $> .05$). Thus, all further analyses were conducted collapsing across sex.

With respect to age, because of the sample's narrow 6-month age range, we did not expect to find strong age-related trends in task performance. Our findings were in line with this prediction, except that age was significantly positively correlated with Time 1 false-belief battery performance (again using the full sample due to the artificially restricted range of the focal sample's scores), $r(30) = .46$, $p = .008$. Because false-belief battery performance is one of our key dependent measures, we elected to control for age in our focal analyses.

Also assessed was children's performance on the PPVT-III (*range*: 27 to 75, $M = 51.71$, $SD = 12.89$) and the pre-false-belief theory-of-mind measure (Diverse Desires: $M = .83$, $SD = .38$; Diverse Beliefs: $M = .63$, $SD = .50$; Knowledge Access: $M = .46$, $SD = .51$; Pre-False-Belief Theory-of-Mind Measure Total: $M = 1.92$, $SD = 0.72$). Performance on the PPVT-III was significantly related to false-belief training task scores, $r(22) = .43$, $p = .038$, and RC-EF aggregate scores, $r(22) = .50$, $p = .013$, and marginally significantly related to Time 2 false-belief battery performance, $r(22) = .37$, $p = .072$. Thus, we controlled for PPVT-III in our focal analyses.

Contrary to our expectations, children's performance on the pre-false-belief theory-of-mind measure was not significantly related to our measures of false-belief improvement (rs $= .03$ to $.20$). One possible explanation for this finding comes from previous work on this scale; according to prior research, children acquire an understanding of diverse beliefs and diverse desires at around 2 years of age. Thus, while tasks measuring these concepts are included in Wellman and Liu's (2004) scale, 3-year-olds would be expected to have already acquired the knowledge necessary to pass these items. In contrast, research suggests that an understanding of both knowledge access and false belief emerges around the 3- to 4-year mark. When looking at the knowledge access task, we found a compelling pattern—in comparison to failers, knowledge access passers had higher false-belief training scores, $t(22) = 2.111$, $p = .046$, and false-belief explanation scores, $t(18) = 3.513$, $p = .002$, as well as more advanced language abilities, $t(22) =$

3.44, $p = .002$, and RC-EF skills, $t(22) = 4.477$, $p < .001$. Thus, for both theoretical and statistical reasons, we created a new variable measuring initial theory-of-mind knowledge that included the four initial false-belief battery tasks, as well as the knowledge access measure ($\alpha = .71$).

Relations Between RC-EF and False-Belief Improvement Measures

Our main question was whether false-belief battery improvement varied with children's performance on the three-item RC-EF aggregate. In order to maximize power while still exploring the potential effects of a number of relevant control variables, we conducted stepwise regressions for the majority of these focal analyses. Our first stepwise regression examined variance in final false-belief battery scores accounted for by RC-EF abilities, age, and scores on both the initial theory-of-mind measure and PPVT-III. This analysis revealed a significant model, $F(1, 22) = 8.68$, $p = .007$, $R^2 = .28$, in which RC-EF performance was the only variable to predict unique variance in final false-belief battery scores ($\beta = .532$, $t = 2.95$, $r_{\text{partial}} = .532$, $p = .007$). Thus, RC-EF was a unique predictor of the extent to which children's performance on the false-belief battery improved over the course of the study.

We then conducted several analyses to address the question of whether stronger RC-EF predicted children's false-belief improvement during the training period. Our first analysis looked at whether RC-EF aggregate scores explained unique variance in how many training tasks children responded to correctly. A stepwise regression including RC-EF aggregate scores, age, initial theory-of-mind scores, and PPVT-III scores revealed a significant model, $F(1, 22) = 13.71$, $p = .001$, $R^2 = .38$. Again, RC-EF performance was the only variable to predict unique variance in training scores ($\beta = .62$, $t = 3.70$, $r_{\text{partial}} = .620$, $p = .001$).

Next, we inspected individual children's performance across the training sessions. Figure 1 shows individual children's performance in each of the training sessions, and a graph showing group performance, grouped by RC-EF aggregate performance (median split). From this figure, it is clear that on the first training task, children uniformly failed. However, during the rest of the first training session, children with stronger RC-EF skills showed much better performance than did children with weaker RC-EF, $t(22) = 2.36$, $p = .029$, $d = 0.96$. These findings suggest that children with stronger RC-EF not only benefited more from the training regimen but that those benefits were realized quickly.

We then examined whether RC-EF aggregate scores were related to the false-belief explanations provided during the second training session. Across all subjects there were 50 correct answers that could be justified—31 from 11 out of 12 children in the high RC-EF group, and 19 from 9 out of 12 children in the low RC-EF group. Figure 2 shows the mean proportion of responses coded into each category by high and low RC-EF groups (median split). As can be seen from inspection of the graph, justifications of children in the high RC-EF group most commonly included reference to relevant factors (e.g., he'll look there "cause that's where it was when he left"). In contrast, children in the low RC-EF group most commonly justified their correct responses by stating false-belief irrelevant facts (e.g., he'll look there because "it's not there").

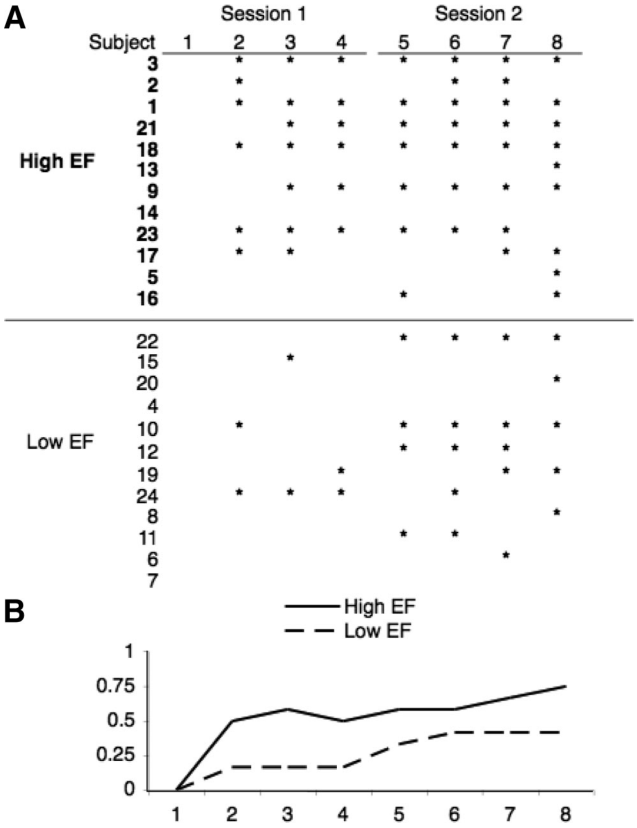

Figure 1. A: Individual children's false-belief training performance by trial (asterisks denote correct performance), listed in order of their response conflict executive functioning aggregate scores (highest at top, division between high and low response conflict executive functioning marking the median split). B: Proportion of all children correctly responding to test questions in each training trial by response conflict executive functioning group. EF = executive functioning.

To account for the fact that some children provided more than one explanation (depending on how many test questions they responded correctly to), each of the 20 children who gave one or more correct false-belief task explanations received a score that reflected the proportion of false-belief relevant explanations provided. We performed a stepwise regression analysis measuring the unique effects of our variables of interest on these scores. This analysis revealed that both RC-EF ($\beta = .42$, $t = 2.27$, $r_{\text{partial}} = .482$, $p = .037$) and initial theory of mind ($\beta = .43$, $t = 2.28$, $r_{\text{partial}} = .484$, $p = .036$) were significant, unique predictors of false-belief explanation scores, $F(1, 17) = 8.12$, $p = .003$, $R^2 = .49$. These findings suggest that both initial theory-of-mind knowledge and RC-EF were positively associated with learning about how the characters' past experiences play a critical role in predicting characters' future actions.

Controlling for RC-EF Improvement

A potential confound in attributing children's false-belief improvement to learning from training was the possibility that the training administered may have inadvertently caused significant improvements in children's RC-EF abilities. To test

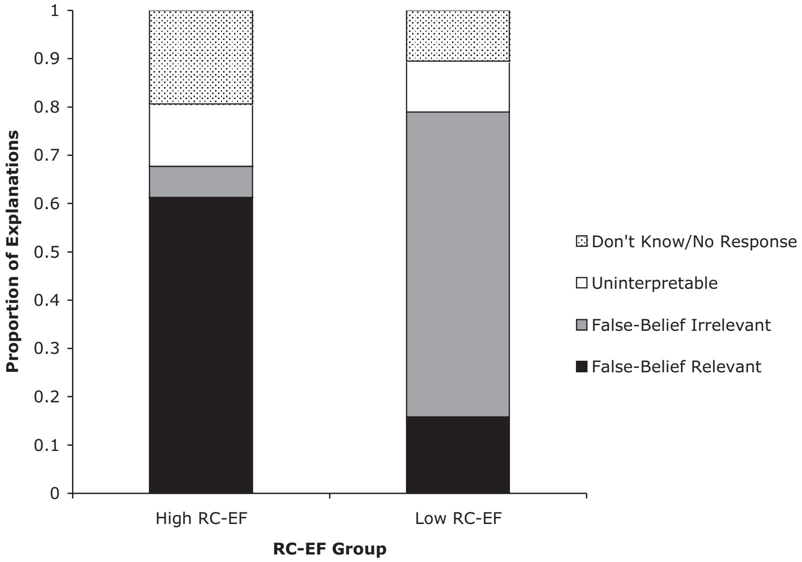

Figure 2. Proportion of explanation types provided by high and low response conflict executive functioning (RC-EF) groups respectively on false-belief training tasks.

this possibility directly, we calculated an RC-EF change score for each child by computing the difference between initial and final practice and test scores on the Bear/Dragon and Grass/Snow tasks independently. The four resulting improvement scores were then standardized and combined to form an RC-EF improvement aggregate score ($\alpha = .63$). To test the extent to which this measure predicted false-belief performance, we added the RC-EF change score to each of the foregoing stepwise regression analyses. In each case, we found that adding RC-EF improvement as a predictor variable to the stepwise regressions led to no changes in the final models reported above. Thus, children's false-belief improvement seems unlikely to be explained by concomitant RC-EF improvement.

Distraction During Training

One possible reason that children with high RC-EF skills were better able to benefit from training was that they were able to maintain their attention during the training sessions and, as a consequence, were better able to encode the relevant feedback (Garon et al., 2008; Posner & Rothbart, 1998; Rothbart & Posner, 2001). To assess this possibility, children's behavior during the administration of the false-belief training tasks was coded for 3 indices of distraction: (a) irrelevant speech, (b) touching or grabbing the stimuli, and (c) looking away for 3 or more consecutive seconds. Four of the 24 children showed no signs of distraction. The average frequency of the three forms of distraction across the study sample was 2.25 for irrelevant speech, 1.63 for stimulus touching, and .25 for looking away for 3 or more consecutive seconds. Each child's standard scores on the three distraction measures were aggregated to create a single measure ($\alpha = .63$). Overall, the distraction aggregate was significantly, inversely related to RC-EF aggregate scores, $r(22) = -.414$, $p = .044$, indicating that indeed, higher levels of distraction during the false-belief training tasks were associated with poorer performance on the RC-EF tasks.

To test whether distraction could account for the relation between RC-EF and false-belief improvement, we reran the stepwise regressions reported above, adding the distraction aggregate to the list of predictor variables. The distraction measure explained significant variance only in false-belief training scores ($\beta = .37$, $t = 2.206$, $r_{partial} = .434$, $p = .039$), but to our surprise, higher levels of distraction were associated with *better* performance on the false-belief training tasks. It is unclear to us why such a relation would exist. Most important, however, is that RC-EF remained a significant predictor of unique variance in all three models examining false-belief improvement (final false-belief battery: $\beta = .53$, $t = 2.95$, $r_{\text{partial}} = .532$, $p = .007$; false-belief training: $\beta = .78$, $t = 4.57$, $r_{\text{partial}} = .706$, $p < .001$; false-belief explanations: $\beta = .42$, $t = 2.27$, $r_{\text{partial}} = .482$, $p = .037$). Thus, while distractibility during false-belief training was negatively correlated with RC-EF skills, there was no evidence that distractibility accounted for the relation between false-belief improvement and RC-EF skills.

Discussion

In this study, we demonstrated that individual differences in preschoolers' RC-EF skills predicted the extent to which their initially poor false-belief performance improved following training. This finding was manifest in three ways: posttest data, improvement during training, and in how children justified their performance during training. The relation between RC-EF and RTM improvement remained strong and significant after accounting for relevant covariates (i.e., age, receptive language ability, initial theory-of-mind knowledge, RC-EF improvement, and distractibility during training).

This pattern of results is consistent with the hypothesis that RC-EF facilitates RTM development by enhancing children's ability to learn from relevant experiences. In the next sections, we raise three possible caveats to this strong claim and discuss the relevant evidence. We then briefly discuss possible mechanisms through which RC-EF may affect learning about false beliefs. Finally, we consider more general implications of our findings for understanding false-belief development in the preschool years.

Did Children Truly Benefit From Training?

An important assumption underlying our claim that individual differences in RC-EF predict children's ability to benefit from training is that the advances we observed in children's false-belief performance were indeed caused by our training regimen. Although we did not include a control group in our study to explicitly test this question, several findings from the extant research support this assumption. First, the research on which we modeled our training paradigm showed similar rates of improvement on the location change task in comparison to what we saw in our own training (Hale & Tager-Flusberg, 2003). More important, those researchers included a control condition in which children received training on relative clauses, and children in that condition did not show similar improvements in false-belief performance over the short testing period. These findings are similar to the control conditions from several other training studies (e.g., Slaughter, 1998; Slaughter & Gopnik, 1996). Also, the striking trajectory of improvement we observed over the course of the training period is not typical of that seen in general false-belief development. False-belief development is usually slow—microgenetic studies show that when children are given no training, individual children show meager improvements from 3.5 to 4 years of age (e.g., Amsterlaw & Wellman, 2006; Flynn, 2006). In contrast, we saw dramatic improvements, especially in the high RC-EF group, over the course of just 2 weeks. This lends confidence that our training regimen was indeed responsible for the sizeable improvement noted in false-belief task performance in this sample.

What Exactly Did Children Learn From Training?

From the outset, our study was focused on the acquisition of one entailment of false-belief understanding, namely that individuals will search for an object based on where they last saw it. We believe that our high RC-EF participants learned something relevant to this principle, as their false-belief explanations reflected this knowledge. Children in the high RC-EF group commonly provided response explanations that could be construed as relevant explanatory rules (e.g., he'll look there "cause that's where he left it when he leaved."), while children in the low RC-EF group often provided irrelevant facts or enunciated far more mechanistic rules (e.g., he'll look there because "it's not there"). These findings suggest that children with advanced RC-EF skills were better able to use their training experience to extract and use relevant information that helped them predict behavior during subsequent tasks.

It is also possible that differences in correct false-belief explanations reflected differences in *conscious awareness* of having used a new predictive strategy, rather than *qualitative differences* in the strategies gleaned from training. Individual differences in RC-EF have been linked to differences in the likelihood of conscious reflection on one's knowledge (Zelazo et al., 2008). Shifts from incorrect to correct responding on our training tasks may therefore have been driven by the same types of changes in underlying predictive strategies across our sample, with the low RC-EF children simply lacking conscious insight into the novel strategies they themselves were employing in instances where they showed improvement. Indeed, microgenetic research on the acquisition of mathematical strategies suggests that adopting a new strategic approach to solve a problem is not always accompanied by conscious recognition of the newly employed strategy (Siegler & Jenkins, 1989). Intriguingly, this research also suggests that children's conscious awareness of having used a new strategy is correlated with the rate at which they subsequently generalize the strategy to other situations (Siegler & Jenkins, 1989). If our false-belief explanation data reflect differences in conscious awareness of novel strategies, then the positive association between RC-EF and explanation may be attributed to the role that RC-EF plays in generalizing the information gleaned from experience with training to other situations.

A more open question is whether RC-EF skills promote more generalized false-belief-relevant learning. Given that our evidence on this question was mixed, it is helpful to consider what generalization, as evidenced by transfer to our structurally different false-belief tasks, would require. For transfer to have occurred, children would have had to appreciate that false beliefs and consequent changes in behavior can not only result from outdated visual information (as in the training tasks) but also from both intentionally deceptive cues (as in the deceptive pointing task) and unintentionally deceptive visual information (as in the contents change and appearance-reality tasks). They would additionally have had to recognize that false-belief inducing experiences affect not only search patterns (as in the training and deceptive pointing tasks) but also one's predictions (as in the contents change task) and object assessments (as in the appearance-reality task). Thus, children would have had to extract relatively general false-belief relevant principles from a circumscribed training experience and extend these underlying principles to a number of structurally diverse scenarios.

We did find some limited evidence for transfer—although gains were most notably pronounced for the location-change task (which was the focus of training), some transfer was seen for the deceptive pointing task. With respect to surface characteristics, deceptive pointing is arguably the task most similar to the location change training tasks in that it involved a character searching for an object, and correct responses required children to point to an empty location. In a sense, this pattern of transfer is consistent with patterns shown in other areas of learning; children's learning often takes place in a more task-specific way before generalizing to structurally diverse but conceptually similar tasks (Diamond et al., 2007; Lohmann & Tomasello, 2003). These findings, along with research on analogical reasoning, suggest that maintaining a new explanatory approach is considerably easier for children on problems that are structurally similar to those on which the approach was discovered than on superficially dissimilar problems (Chen, 1996; Goswami, 1992, 1995). Knowledge transfer on the deception task, which was arguably most structurally similar to the false-belief training tasks, may reflect this general characteristic of learning. Similarly, it is noteworthy that recent efforts to train RC-EF skills in children also have tended to show only narrow transfer effects (Diamond & Lee, 2011).

An important question for future research concerns the conditions and factors necessary for further-generalized conceptual transitions to occur; it seems likely that these transitions would rely on more diverse false-belief-relevant experiences over time (see e.g., Amsterlaw & Wellman, 2006). Along these lines, we should clarify that our intention is not to argue that children necessarily naturalistically learn the precise entailment of false belief that we taught them, nor that doing so comes at a particular stage in false-belief development. Moreover, we do not intend to suggest that RC-EF skills are important *only* for false-belief knowledge acquisition; on the contrary, we would expect RC-EF skills to play a significant role in learning across many learning contexts—particularly those that require inhibiting or ignoring salient concepts (e.g., prior incorrect theories or incongruent perspectives). Our findings suggest that domain general RC-EF skills predict the extent to which children benefit from false-belief-relevant feedback in the service of advancing their underlying knowledge. We speculate that the effects shown here would also be shown when considering more naturalistic sources of information about false beliefs, such as those noted from the outset (i.e., parent-child conversation, sibling conversation, shared book reading, etc.).

Can a Case Be Made for the Expression Account?

Our hypothesis regarding the role that RC-EF plays in constructing an understanding of false beliefs contrasts with the alternate expression account of the relation between RC-EF and false belief. Recall that according to the expression account, 3-year-olds fail false-belief tasks primarily because they have deficient RC-EF skills and therefore cannot demonstrate their underlying RTM knowledge on standard measures. Evidence from the present data that this is true would have come if the observed improvements in false belief were tied to improvements in RC-EF. We did not find evidence in support of this account—there was little improvement in RC-EF over the course of the study, and what little improvement that did exist was not related to false-belief improvements.

Still, a different way to offer an expression account is to suggest that practice with false-belief tasks in the training sessions led to a decrease in the executive demands inherent to the tasks. If this were true, then false-belief improvement in children with well-developed RC-EF skills may have occurred because their preexisting RC-EF skills were sufficient to overcome the gradually lowered executive task demands. In contrast, lowered executive task demands from practice with the tasks may not have been sufficient to improve the performance of children with less-developed RC-EF skills to the same degree. Although this is an intriguing possibility, we do not believe it can account for our training findings. First, practice effects were realized quickly, and it does not seem that one trial would be sufficient to substantially reduce the RC-EF demands inherent to the tasks. Second, there is no a priori reason to believe that false-belief explanation tasks have especially strong RC-EF requirements (e.g., Perner, Lang, & Kloo, 2002), yet the benefits of RC-EF skills were realized here as well. Thus, we believe it more likely that RC-EF affected children's ability to learn something relevant from the training, which in turn enabled them to pass the false-belief tasks.

What Role Is RC-EF Playing?

From the outset, we noted that within the theory-theory framework, preschoolers' conceptual understandings of others' mental states are thought to develop through meaningful interactions with others. We believe that our findings can help elucidate how domain-general cognitive skills (i.e., RC-EF) facilitate children's abilities to change theories over time based on experience.

With respect to specific processes, one possibility is that having the ability to inhibit competing thoughts and distractions enabled children to more easily consider and attend to relevant variables. Although our analyses showed that signs of distraction did not account for the relation between RC-EF and false-belief improvement, RC-EF may still have affected the extent to which children noticed or encoded the relevant aspects of our feedback. Even when children are attending to events in their environment, RC-EF skills may be necessary for them to absorb certain types of information. Indeed, several authors have suggested that children may have particular difficulty attending to the mental states of others because doing so requires the suppression of their own mental states (Carlson & Moses, 2001). Similarly, recognition of false beliefs in others may require the capacity to separate or "distance" mental representations from the objective reality they represent (Moses & Sabbagh, 2007). These characteristics of reasoning about subjective mental states may have impacted the extent that low RC-EF children were able to consider and attend to relevant variables when reasoning about false-belief-based behavior.

A second possibility is that children with high RC-EF skills were more effective at noticing discrepancies between expectations and subsequent outcomes. That is, children with more advanced RC-EF skills may have been better able to disengage from their current knowledge states to explicitly reflect on the discrepancy between their previous conceptions and the notions being presented in training (Zelazo, 2004).

Finally, RC-EF skills may have impacted children's ability to integrate false-belief-consistent evidence with prior knowledge. After observing that people usually act in accordance with *true* beliefs, children may come to develop true belief expectations that lead to erroneous predictions in the relatively rare instances where *false* beliefs motivate behavior. If young children come to expect that others' beliefs and subsequent actions will be in accordance with reality, then any advances in their knowledge of false beliefs would necessitate updating incomplete or inaccurate prior belief knowledge. Indeed, the process of integrating new, incongruent evidence with previously established knowledge appears to be particularly challenging for preschoolers (e.g., Siegler & Chen, 1998). Thus, the preschool children in our study may have been reliant on their developing RC-EF skills when facing the task of integrating information from false-belief-based feedback with preexisting mental state knowledge.

Limitations and Future Directions

There are several limitations of this study. For instance, our relatively small sample size may limit the generalizability of our findings. Small sample sizes are a common feature of studies that attempt to get fine-grained detail of developmental phenomenon. In this study, as in others of this nature, replication is therefore critical. A second limitation is that we have examined the effects of RC-EF on a relatively narrow aspect of theory-of-mind understanding that emerges during a circumscribed stage of development. An important question for future research is whether RC-EF plays a role in theory-of-mind development across a range of ages.

Future research is also necessary to directly test whether our findings reflect what occurs in real-world settings. Does RC-EF predict the extent to which children benefit from everyday experiential factors that are known to influence false-belief knowledge, such as parental use of mental state terms? Do our findings reflect how RC-EF and experience interact to contribute to more natural, extended developmental trajectories of false-belief knowledge? Answering these questions would provide more conclusive evidence on the role that RC-EF plays in social-cognitive development.

Other important avenues for future research involve examining the neurological changes that underlie theory-of-mind development and how these changes might themselves be related to both relevant experiential factors and RC-EF skills. The impact of experience on brain structures has been observed in a number of carefully designed studies (e.g., Maguire et al., 2003). However, no research has directly linked the development of brain regions associated with RTM to experience or to the neuropsychological changes that underlie advances in RC-EF. While our study represents a first step in understanding how RC-EF and experience might work together to foster advances in RTM knowledge, future research should aim to better understand how these factors interrelate on a neurological level.

Conclusion

Our findings show that RC-EF skills predict the extent to which children improve on false-belief tasks following training. This pattern of findings lends preliminary support to the hypothesis that RC-EF plays a role in promoting children's ability to construct an understanding of false belief from relevant experiences. These findings help provide an integrated approach to understanding how experience and RC-EF work together to shape the timetable of RTM development during the preschool years.

References

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7,* 139–172. doi:10.1207/s15327647jcd0702_1

Bartsch, K. (1998). False belief prediction and explanation: Which develops first and why it matters. *International Journal of Behavioral Development, 22,* 423–428. doi:10.1080/016502598384450

Benson, J. E., & Sabbagh, M. A. (2009). Theory of mind and executive functioning: A developmental neuropsychological approach. In P. D. Zelazo, M. Chandler, & E. Crone (Eds.), *Developmental social cognitive neuroscience* (pp. 63–80). New York, NY: Psychology Press.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78,* 647–663. doi:10.1111/j.1467-8624.2007.01019.x

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273–293. doi:10.1207/S15326942DN1903_3

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28,* 595–616. doi:10.1207/s15326942dn2802_3

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: 0009–3920/2001/7204–0007.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69,* 672–691. doi:10.1111/1467-8624.00333

Carlson, S. M., Zelazo, P. D., & Faja, S. (in press). Executive function. In P. D. Zelazo (Ed.), *Handbook of developmental psychology*. New York, NY: Oxford University Press.

Chen, Z. (1996). Children's analogical problem solving: The effects of superficial, structural, and procedural similarity. *Journal of Experimental Child Psychology, 62,* 410–431. doi:10.1006/jecp.1996.0037

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:10.1111/1467-8624.00061

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318,* 1387–1388. doi:10.1126/science.1151148

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333,* 959–964. doi:10.1126/science.1204529

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition*. Bloomington, MN: Pearson Assessments.

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26,* 465–486. doi:10.1207/s15326942dn2601_6

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,* i, iii, v, 1–87. doi:10.2307/1165866

Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology, 24,* 631–654. doi:10.1348/026151005X57422

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science, 7,* 103–115. doi:10.1111/j.1467-7687.2004.00326.x

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

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134,* 31–60. doi:10.1037/0033-2909.134.1.31

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

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld & S. Gelman (Eds.), *Domain specificity in cognition and culture* (pp. 257–293). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511752902.011

Goswami, U. (1992). *Analogical reasoning in children*. Hillsdale, NJ: Erlbaum.

Goswami, U. (1995). Transitive relational mappings in three- and four-year-olds: The analogy of *Goldilocks and the Three Bears*. *Child Development, 66,* 877–892. doi:10.2307/1131956

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6,* 346–359. doi:10.1111/1467-7687.00289

Henning, A., Spinath, F. M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology, 108,* 513–531. doi:10.1016/j.jecp.2010.10.006

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16,* 233–253. doi:10.1111/j.2044-835X.1998.tb00921.x

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34,* 1326–1339. doi:10.1037/0012-1649.34.6.1326

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34,* 1026–1037. doi:10.1037/0012-1649.34.5.1026

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology, 28,* 645–668. doi:10.1207/s15326942dn2802_5

Kochanska, G., Murray, K., Jacques, T. Y., & Koenig, A. L. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67,* 490–507. doi:10.2307/1131828

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1,* 247–253. doi:10.1111/1467-7687.00038

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

Maguire, E. A., Spiers, H. J., Good, C. D., Hartley, T., Frackowiak, R. S. J., & Burgess, N. (2003). Navigation expertise and the human hippocampus: A structural brain imaging analysis. *Hippocampus, 13,* 250–259. doi:10.1002/hipo.10087

Moses, L. J., & Sabbagh, M. A. (2007). Interactions between domain general and domain specific processes in the development of children's theories of mind. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 275–291). New York, NY: Psychology Press.

Oh, S., & Lewis, C. (2008). Korean preschoolers; advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development, 79,* 80–99. doi:10.1111/j.1467-8624.2007.01112.x

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. doi:10.1126/science.1107621

Pears, K. C., & Moses, L. J. (2003). Demographics, parenting and theory of mind in preschool children. *Social Development, 12,* 1–20. doi:10.1111/1467-9507.00219

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology, 43,* 974–990. doi:10.1037/0012-1649.43.4.974

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. doi:10.1111/1467-8624.00436

Perner, J., Leekam, 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. doi:10.1111/j.2044-835X.1987.tb01048.x

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society: Series B*. Biological Sciences, *353,* 1915–1927. doi:10.1098/rstb.1998.0344

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

Rothbart, M., & Posner, M. (2001). Mechanism and variation in the development of attentional networks. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 353–363). Cambridge, MA: MIT Press.

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (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., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73,* 734–751. doi:10.1111/1467-8624.00435

Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (2009). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development, 80,* 1147–1162. doi:10.1111/j.1467-8624.2009.01322.x

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81. doi:10.1111/j.1467-9280.2005.01667.x

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology, 36,* 273–310. doi:10.1006/cogp.1998.0686

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development, 69,* 321–332. doi:10.1111/j.1467-8624.1998.tb06191.x

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67,* 2967–2988. doi:10.2307/1131762

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. doi:10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523–541. doi:10.1111/j.1467-8624.2004.00691.x

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

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences, 8,* 12–17. doi:10.1016/j.tics.2003.11.001

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297–301. doi:10.1038/nprot.2006.46

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). Development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Appendix

Sample False Belief Training Task Script

Place a miniature bush, a miniature log, a rabbit, and two characters on the table in front of the child. “This girl’s name is Sonia, and this is her brother Eddie.” Show the child the two dolls. “Sonia and Eddie have a pet rabbit. They like to play with the rabbit in their back yard. One day while they’re playing, their rabbit hops into the bush.” Show the rabbit moving into the bush until it is out of sight. “Then Eddie goes inside to get a snack.” Move Eddie off the table, out of sight. “While he’s inside, Sonia wants to play with the rabbit some more, so she gets the rabbit out of the bush.” Show Sonia retrieving the rabbit from the bush. “And then they play some more. When she’s done, Sonia puts the rabbit away, but she puts it in the log.” Show Sonia placing the rabbit inside the log, out of sight. “Then Sonia goes away to play in the park.” Move Sonia off the table, out of sight. “A little while later, Eddie comes back outside and he wants to play with the rabbit some more.

So, [child’s name], where will Eddie look for the rabbit?” Wait for response. During the second testing session, the experimenter additionally asks, “Why will he look there?”

If the child chooses the correct location, provide the following feedback: “That’s right. Eddie *does* look in the bush for his rabbit.” Show Eddie looking in the bush for his rabbit. “Eddie was gone when Sonia took the rabbit and put it in the log, so Eddie didn’t see the rabbit get moved! When Eddie comes back, he’ll look in the bush for his rabbit, ‘cause that’s where it was when he left!”

If the child chooses the incorrect location, provide the following feedback: “Well, that’s a really good guess, but that’s not right! When Eddie comes back, he’ll look in the bush for his rabbit. Let’s watch again and see why.” Place all the objects in their original positions. “Eddie saw the rabbit go into the bush.” Place the rabbit back in the bush. “Then he left!” Move Eddie off the table, out of sight. “While he was gone, Sonia got the rabbit and put it in the log.” Show Sonia retrieving the rabbit, and placing it in the log. “Eddie was gone, so he didn’t see that! He didn’t *see* Sonia put the rabbit in the log. So, when Eddie comes back, he’ll look in the bush for the rabbit, ‘cause that’s where it was when he left!” Show Eddie looking in the bush to retrieve the rabbit.

Received June 9, 2010
Revision received July 30, 2012
Accepted August 8, 2012 ■