

Executive Functioning and Preschoolers' Understanding of False Beliefs, False Photographs, and False Signs

Mark A. Sabbagh
Queen's University at Kingston

Louis J. Moses
University of Oregon

Sean Shiverick
University of Wisconsin at Madison

Two studies were conducted to investigate the specificity of the relationship between preschoolers' emerging executive functioning skills and false belief understanding. Study 1 ($N = 44$) showed that 3- to 5-year-olds' performance on an executive functioning task that required selective suppression of actions predicted performance on false belief tasks, but not on false photograph tasks. Study 2 ($N = 54$) replicated the finding from Study 1 and showed that performance on the executive functioning task also predicted 3- to 5-year-olds' performance on false sign tasks. These findings show that executive functioning is required to reason only about representations that are intended to reflect a true state of affairs. Results are discussed with respect to theories of preschoolers' theory-of-mind development.

Theory-of-mind research has become a central focus of research in developmental psychology over the past 20 years, largely because of its far-reaching theoretical and empirical implications for various branches of the discipline, including cognitive development, social development, and child clinical psychology (Carpendale & Lewis, 2005; Repacholi & Slaughter, 2003). The most concentrated focus has been on a transition during the preschool period in children's reasoning about false beliefs (Wellman, Cross, & Watson, 2001). At the beginning of this period, children appear to be unable to report that they or others can believe something that is not true. This ability to reason correctly about false beliefs typically emerges by the time they are 4 or 5 years of age. A recent meta-analysis has shown that this age effect is quite robust across a wide variety of task manipulations (Wellman et al., 2001).

A prevalent view is that children who fail tasks designed to tap false belief reasoning lack certain insights into the nature of beliefs themselves (Wellman et al., 2001). For instance, Bartsch and Wellman (1994) argue that children may not understand the

role that beliefs play in motivating others' actions, opting instead to give weight to other mental constructs, such as desires or intentions. In a similar vein, Perner (1991b) suggests that children may not understand that beliefs are limited representations of the world, related to but ultimately distinct from the reality they are meant to represent. In either case, theorists within this framework posit that a fundamental conceptual advance is required to pass the false belief task and other preschool theory-of-mind tasks such as those tapping children's understanding of deception, perceptual perspective taking, and appearance-reality (see Taylor, 1996, for a review).

That said, factors other than conceptual changes may also be implicated in the transition to false belief understanding. One such factor is executive functioning (Carlson, Moses, & Hix, 1998; Frye, Zelazo, & Palfai, 1995; Russell, 1996). Executive functioning skills include domain-general cognitive abilities such as planning, working memory, action monitoring, and inhibitory control. Such skills are clearly necessary for successful performance on false belief tasks. For example, a typical false belief task requires (1) cognitive inhibition to disengage from a salient real-world situation to attend to an intangible abstract representation, (2) response inhibition to inhibit a prepotent or habitual way of responding (i.e., pointing to the "true" location of the object), and (3) working memory to indicate the correct answer while holding in mind two different and conflicting

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Mark Sabbagh. We thank Samantha Samonte, Laura Coristine, and Stephanie Jull for their assistance with data collection and coding. Thanks are also due to the children and parents who volunteered their time to participate in these studies.

Correspondence concerning this article should be addressed to Mark Sabbagh, Department of Psychology, Queen's University, Kingston, ON, Canada K7L 3N6. Electronic mail may be sent to sabbagh@post.queensu.ca.

© 2006 by the Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2006/7704-0014

representations (i.e., their own and that of the protagonist). Intriguingly, a number of studies have now demonstrated a sizable relation between theory-of-mind and executive functioning (Carlson & Moses, 2001; Davis & Pratt, 1995; Frye et al., 1995; Hala, Hug, & Henderson, 2003; Hughes, 1998; Perner & Lang, 1999).

One technique that has been used to assess whether domain-general as well as domain-specific developments are important for false belief and theory-of-mind involves comparing performance on false belief and so-called "false" photograph tasks (Zaitchik, 1990). The false photograph task is designed to parallel closely the false belief task in terms of formal structure and processing demands. The major difference is that children are required to make judgments about photographic representations that do not accurately reflect current reality rather than about mental representations that do not do so. In false photograph tasks, children listen to a story in which a character puts an object in one place and then takes a Polaroid photograph depicting the object in that place. While the photograph is developing, the object is moved to a new location, thereby rendering the photograph inaccurate regarding the location of the object. As in the false belief task, children are asked to report the contents of the false representation (i.e., "In the picture, where is [the object]?"). Consistent with these formal similarities, a number of studies have shown that performance on the two tasks shares a similar developmental trajectory—3-year-olds consistently fail both tasks and systematically correct performance emerges sometime around the age of 4 or 5 (Davis & Pratt, 1995; Leekam & Perner, 1991; Leslie & Thaiss, 1992; Zaitchik, 1990).

Because of the similarities in task structure, it has been argued that reasoning in both tasks requires equivalent domain-general skills (Leslie & Thaiss, 1992). Both tasks require children to inhibit a prepotent tendency to refer to a true location in order to refer to a false one. Also, both tasks require children to hold in mind two competing representations of the same situation. The executive demands of the tasks are thus apparently quite similar. Hence, it is argued that any behavioral dissociations between these two tasks can be most readily attributed to differences in domain-specific knowledge. As a result, comparing the developmental profiles of performance on false belief and false photograph tasks is often seen as a critical test for assessing the relative weights of domain-specific and domain-general processes in theory-of-mind development.

Indeed, performance on these two tasks can be dissociated in a number of ways. First, despite their apparent similarities, performance on the false belief and false photograph tasks is only weakly correlated (e.g., Davis & Pratt, 1995; Slaughter, 1998). Second, training on one does not generate improved performance in the other (Slaughter, 1998). Third, and perhaps most strikingly, individuals with autism are impaired on false belief tasks but not on false photograph or false drawing tasks (Charman & Baron-Cohen, 1992; Leekam & Perner, 1991; Leslie & Thaiss, 1992). Finally, in adults, the neural systems activated during false belief reasoning are distinct from those associated with false photograph reasoning (Sabagh & Taylor, 2000). These findings clearly show that, despite apparent structural similarities, false belief and false photograph reasoning have quite different ontogenetic and neurocognitive profiles.

That said, just how to characterize dissociations between false belief and false photograph reasoning is not clear. One possibility, as suggested earlier, is that performance on false belief and false photograph tasks differs because of domain-specific knowledge requirements. Accordingly, ontogenetic and neurocognitive dissociations can be seen as strictly related to differences in conceptual acquisition and representation. An alternative possibility, however, is that despite the surface similarity of belief and photograph tasks, reasoning about representations in different domains may nonetheless pose differential demands on domain-general cognitive processes. That is, the extent to which domain-general skills such as executive functioning are required may depend on the nature of the representation. One way to tease apart these two possibilities is to examine whether individual differences in executive functioning relate to children's performance on the false photograph task as strongly as they do to their performance on the false belief task. If they do, this would support a strong version of the claim that domain-specific conceptual developments are the primary engine of theory-of-mind development in the preschool period. If they do not, it would suggest that there are differences in the extent to which representations in different domains rely on executive functioning skills, and that domain-general advances may well be critical to theory-of-mind development in addition to conceptual advances.

There are at least two reasons to think that false belief and false photograph tasks may pose differential executive demands. First, Russell, Saltmarsh, and Hill (1999) have suggested that while mental states are clearly abstract and, as such, potentially less salient than reality itself, it is not clear that the

same is true of photographs. The physical instantiation of the photograph itself may lessen the executive demand of “disengaging from reality” in order to reason about the representation. Thus, judgments about photographs may rely less on executive functions than do judgments about beliefs.

Second, functional differences between beliefs and photographs themselves may generate differential executive demands. The “job description” of a belief is to be an accurate representation of some real state of affairs. Importantly, we attempt to maintain the veracity of a belief over time by updating it when the world changes (so long as we are aware of the change). In contrast, photographs have a very different job description. Photographs are typically intended to correspond with the world the moment they are taken, but not after. To illustrate, there is nothing anomalous or problematic about a photograph depicting a friend skiing in Aspen when in fact that friend is currently sitting on the couch in the middle of summer. However, believing that the same friend is in Aspen under those same circumstances would obviously be undesirable. This difference in the function of beliefs and photographs may generate an imbalance in the extent to which tasks assessing understanding of these representations tax the executive system. In order to think about a false belief, children need to suppress a potentially prepotent assumption that beliefs are faithful up-to-date representations of reality. This same need is not present when children reason about photographs, and hence the executive demands of false photograph tasks might be relatively light.

In summary, a comparison of the false belief and false photograph tasks may open a unique window on the cognitive and conceptual developments necessary for the emergence of theory-of-mind reasoning in the preschool years. To date, there is strong evidence to suggest that false belief and false photograph performance are dissociated in development. However, no research has directly addressed the source of this dissociation. In the following two studies, we explored this critical issue.

Study 1

The goal of our first study was to determine whether individual differences in executive functioning correlate with performance on both the false belief and false photographs tasks. To this end, we tested children on false belief and false photograph tasks as well as three executive functioning tasks: bear/dragon, whisper, and gift delay (Carlson & Moses, 2001; Kochanska, Murray, Jacques, Koenig, & Van-

degeest, 1996). These three executive functioning tasks were chosen because they measure somewhat different aspects of executive functioning. Specifically, the bear/dragon task (described below) measures children’s ability to choose between two conflicting response types (performing or suppressing an action) based upon a rule. The bear/dragon task is thought to require both inhibitory control (in suppressing an action) and working memory (in remembering the rule) (Carlson, Moses, & Breton, 2002; Hala et al., 2003). The whisper task simply requires children to lower the intensity of a vocal response. Finally, the gift delay task merely requires children to delay engaging in a desired activity. While the latter two tasks impose inhibitory demands, they would appear to make only minimal working memory demands (see Carlson et al., 2002). Previous work has shown that all three of these tasks are reliable constituents of a larger battery of executive functioning tasks, and have been used successfully with both Western and Non-Western populations (Carlson & Moses, 2001; Kochanska et al., 1996; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Moreover, these tasks tend to correlate strongly with measures of self-consciousness and reflectivity that would also seem to require executive functioning (Kochanska et al., 1996). Finally, the tasks have been found to consistently relate differentially to theory-of-mind tasks; conflict tasks like bear/dragon typically relate more strongly to theory-of-mind than do the simpler inhibition tasks (Carlson & Moses, 2001; Carlson, Moses, & Claxton, 2004; Carlson et al., 2002). This is arguably because both bear/dragon and theory-of-mind tasks require children to suppress one habitual way of responding while in addition holding in mind the information that impels one to respond otherwise.

Thus, we predict that of the three executive functioning tasks, bear/dragon will have the strongest relation with theory-of-mind. Nonetheless, we included two additional executive tasks to explore whether a differential relation to executive functioning might also be found for reasoning about false photographs.

Method

Participants

Forty-four preschool children (24 female) aged between 3 years 0 months and 5 years 2 months ($M = 3$ years 9 months) participated in the study. For certain analyses, participants were divided into two age groups based on median age split. For simplicity,

we will refer to these as the younger ($M = 3$ years 6 months; range: 3 years 0 months to 3 years 9 months) and older ($M = 4$ years 5 months; range: 3 years 10 months to 5 years 6 months) children. Participants were recruited from a database drawn from a primarily White middle-class community in the Pacific Northwest of the United States. Six additional children participated but did not complete the battery of executive functioning tasks, and therefore were excluded from analysis.

Measures

A single male experimenter administered a small battery of tasks to assess children's performance on false belief, false photograph, and executive functioning tasks. The materials and procedures used for each of these tasks are described below. For each of the false photograph and false belief tasks, children were judged as passing only if they answered control questions and test questions correctly.

False belief: location change. In this adaptation of Wimmer and Perner's (1983) classic task, children were introduced to two dolls (Ernie and Bert from Sesame Street) playing with a ball on a tabletop that also contained a red and blue box (each $23 \times 15 \times 15$ cm). The puppets were briefly shown playing with the ball until the experimenter explained that Bert had become hungry and wanted to eat lunch. Bert then put the ball inside the blue box and left the table. The experimenter then explained that Ernie was not done playing and then portrayed Ernie as retrieving the ball and playing with it by himself. After a while, Ernie himself got hungry and placed the ball into the red box. Ernie then left the table and Bert returned, wanting to play with his ball. Children were then asked two questions in fixed order: (1) *False Belief Test Question*: "Where does Bert think the ball is?" and (2) *Reality Control Question*: "Where is the ball really?"

False belief: unexpected contents. Following Gopnik and Astington (1988), children were shown a Band-Aid box and asked what they thought was inside. Once children said "Band-Aids" the box was opened to show that it actually contained crayons. The crayons were then replaced, the box was closed, and children were asked the "*Self*" Test Question: "When you first saw this box, before we looked inside there, what did you think was inside? Band-Aids or crayons?" After children responded, a naïve doll was presented and children were asked the "*Other*" Test Question: "Here's Bert! Bert has never seen inside this box before. What does Bert think is in this box? Band-Aids or crayons?" Finally, children were asked

a *Reality Control Question*: "What is really in this box? Band-Aids or crayons?" To allow for maximum commensurability with the false photograph task, we scored only performance on the "other" question.

False photograph: location change. This task was modeled after Zaitchik's (1990) false photograph task, which itself was designed to be roughly comparable with the false belief task. To begin, children were familiarized with a Polaroid camera by taking a photograph of a toy dinosaur. Children took the photograph themselves and then sat with the experimenter as it developed. As the photograph developed, the experimenter pointed out the features of the toy dinosaur in the photograph.

After the camera demonstration, children were introduced to a toy puppet (Ernie) sitting on a chair. The experimenter helped the child to take a picture of Ernie sitting on the chair. The photograph was withdrawn from the camera and placed face down on the table. The experimenter then explained that Ernie was tired and wanted to lie down. Ernie was then moved to a nearby bench. Children were asked two questions in fixed order: (1) *Control Questions*: "When we first took the picture, where was Ernie sitting? In the chair or on the bench?" and "Where is he sitting now? In the chair or on the bench?" and (2) *False Photograph Test Question*: "In the picture [experimenter touches the photograph] where is Ernie? In the chair or on the bench?"

False photograph: identity change. This task was modeled after that of Leslie and Thaiss (1992). In this version, the experimenter first helped the child to take a picture of a bee puppet sitting on a table. After the picture was taken and placed face down on the table, the experimenter enacted a story in which the bee flew away and was replaced by a zebra puppet. The children were then asked two kinds of questions, again in fixed order: (1) *Control Questions*: "Who was sitting on the table when we first took the picture? The zebra or the bee?" and "Who is sitting on the table now?" and (2) *False Photograph Test Question*: "In the photograph [experimenter touches the photograph] who is sitting on the table? The zebra or the bee?"

Whisper task. This measure, adapted from Kochanska et al. (1996), requires children to speak in a whisper while naming familiar cartoon characters. After the experimenter explained (in a whispering voice) to children that the goal of the game was to keep whispering, children were given a practice trial in which they were asked to whisper their name. Children who did not whisper in this warm-up trial were given the instructions again and the trial was

repeated until they did whisper. After eliciting children's names in a whisper, the experimenter showed children pictures of 10 cartoon characters that were printed on laminated cards (10 × 15 cm). The cards were presented one at a time while the experimenter whispered the question "who's this?" There were six relatively familiar cartoon characters (Big Bird, Poohontas, Donald Duck, Snow White, the Beast from "Beauty and the Beast," and Mickey Mouse) and four relatively unfamiliar characters (Huckle, Elmer Fudd, Petunia Pig, and Fat Albert). The order of the cards was random, with the constraint that no more than two familiar or unfamiliar cards were presented consecutively. After the first five trials, children were reminded to maintain the whisper. For each item in which children used the character's name (as opposed to saying "I don't know," or providing no response at all), responses were scored by the following coding scheme: 0 = *a shout*, 1 = *a normal or mixed voice*, and 2 = *a whisper*. The internal reliability of this task was high ($\alpha = .83$) and therefore scores were aggregated across known items. Because children differed in the number of characters they knew (and hence in the number of trials they completed), the dependent variable was the average response score over completed trials (range: 0–2). Interrater reliability was calculated on 25% of the transcripts and found to be of an acceptable level (agreement = 88%, $\kappa = .68$).

Bear/dragon task. This simplified version of a "Simon Says" game was developed originally by Reed, Pien, and Rothbart (1984) and modified more recently by Kochanska et al. (1996). First, children were given a warm-up in which they were asked to perform 10 "silly" gestures (e.g., stick out your tongue, touch your ears, clap your hands, etc.). After ensuring that children could carry out all 10 gestures, the experimenter introduced the bear and dragon puppets. The bear (36 cm tall, brown, furry) was described as "A nice bear, our friend. When the bear asks us to do something, we do what he says." The dragon (25 cm tall, felt, with scary eyes and teeth) was described as "A mean old dragon. We don't listen to him. If he asks us to do something, we don't do it." Following these descriptions were two practice trials. In the bear practice, the experimenter assumed a high-pitched friendly voice and said "Touch your nose." In the dragon practice, the experimenter assumed a low gruff voice and said "Touch your ears." Although no children had difficulty with the bear practice trial, children frequently failed the dragon practice trial (i.e., performed the action even though prohibited by the rules). The dragon practice trial was repeated up to six times

until children performed correctly. Following the practice trials, participants were tested to ensure they understood the rule (i.e., "If the bear asks you to do something, are you going to do it?"). Once children demonstrated that they understood the rule, the 10 test trials followed (5 bear, 5 dragon in alternating order). After the first 5 trials, the experimenter reminded all children of the rules.

For each dragon trial, children were given a score of 3 if they did not carry out the commanded action, 2 if they did not carry out the target action but made some other movement, 1 if they partially carried out the action, and 0 if they clearly carried out the action. Internal reliability of this measure was excellent ($\alpha = .88$), and thus scores were summed across trials and created a range from 0 to 15 possible. In addition, the number of dragon practice trials in training was used as a second dependent measure. Interrater reliability was calculated on 25% of the transcripts and found to be acceptable (92% agreement, $\kappa = .87$).

Gift delay task. In this task developed by Kochanska et al. (1996), children were asked to seat themselves in a chair facing away from a table (toward a video camera) while a present is wrapped behind their backs. Children are enjoined to not turn around and look while the present is being wrapped. The experimenter then noisily wraps the present for 60 s. The videotape of the child was scored on three dimensions: (1) peek resistance (0 = *full torso turn*, 1 = *peek over shoulder*, 2 = *no peeking*), (2) total number of peeks, and (3) time latency to first peek. These measures were highly intercorrelated ($r_s = .82-.84$), thus establishing that they could be aggregated to form a measure with good internal reliability. Interrater reliability was calculated on 25% of the transcripts and found to be acceptable (for resistance, agreement = 73%, $\kappa = .53$; for number, agreement = 91%, $\kappa = .85$; for latency, agreement = 82% within 2 s).

Design

Participants were tested individually in an on-campus playroom at a child-sized table. All tasks were administered in a single session that lasted approximately 30 min. The entire session was videotaped and all measures were scored from video. As is standard practice in individual differences research, the tasks were administered to all subjects in the same fixed order: (1) false photograph: location change, (2) whisper, (3) false belief: location change, (4) bear/dragon, (5) false belief: unexpected contents, (6) gift delay, and (7) False photograph: identity change. A fixed order is used to ensure that any

order effects are roughly similar across individuals, and thus cannot affect the individual differences relations among tasks.

Results

The main goal of the present study was to address the relations between false belief, and false photograph understanding, and three measures of executive functioning. Before presenting the analyses relevant to this question directly, we will briefly describe the results related to each of the tasks individually.

False Photograph and False Belief Tasks

Four children were excluded from the analysis because of a failure to answer control questions correctly. Preliminary analyses revealed no differences in children's performance across the two trials within both the false belief or false photograph tasks. Thus, children's scores were collapsed across trials to create a single false belief and a single false photograph measure (range for each: 0–2). The means and standard deviations for these measures are presented in Table 1.

A 2 (age) \times 2 (task) mixed-design ANOVA with task as a within-subjects factor revealed a significant main effect of task, $F(1, 40) = 13.01$, $p < .05$, $\eta^2 = .25$. Collapsed across ages, children found false belief more difficult ($M = 0.78$, $SE = .10$) than false photograph ($M = 1.26$, $SE = .11$). The analysis also revealed a main effect of age, $F(1, 40) = 14.39$, $p < .05$, $\eta^2 = .27$. Collapsed across tasks, younger children ($M = 0.71$, $SE = .12$) performed more poorly than older children ($M = 1.33$, $SE = .11$).

Finally, performance on false photograph was moderately correlated with performance on false belief, $r(42) = .32$, $p < .05$. However, when general improvements related to age were statistically controlled, this relation was no longer significant, $r(41) = .11$, *ns*.

Executive Function Tasks

The raw means and standard deviations for each of the executive function tasks are shown in Table 1.

Whisper task. Children's score on the whisper task was calculated as their average whisper score for known items to which they gave a character name response (range: 0–2). As expected, the whisper measure was correlated with age, $r(44) = .31$, $p < .05$.

Bear/dragon task. Three children failed to complete this task and therefore were excluded from any

Table 1
Means and Standard Deviations for Tasks (with Range of Scores) Used in Study 1

Age group	<i>n</i>	Executive functioning tasks															
		Representation tasks						Bear/ dragon				Gift delay				Whisper (0–2)	
		False belief (0–2)		False photographs (0–2)		Practice (1–5)		Total (0–15)		Peek resistance (0–2)		Total number of peeks (0–5)		Latency (0–60 s)		<i>M</i>	<i>SD</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Younger	19	0.42	.50	1.00	.82	2.55	1.71	9.86	5.78	1.43	.73	1.17	1.75	43.22	22.13	1.64	.52
Older	23	1.13	.76	1.52	.59	1.14	0.48	14.48	1.36	1.60	.72	0.61	1.20	50.22	18.64	1.77	.28

analyses involving the task. Two dependent measures reflected children's performance in the bear/dragon task: (1) number of dragon practice trials (max. = 6) and (2) score on the dragon trials (out of 15). Preliminary analyses revealed that these two dependent measures were intercorrelated, $r(41) = -.51, p < .05$. Thus, the scores were standardized (practice trials were reverse scored) and summed to form a composite bear/dragon score. Again, this measure of executive functioning was highly correlated with children's age, $r(41) = .56, p < .05$.

Gift delay task. There were three dependent measures that captured performance in the gift delay task: (1) peek resistance (range: 0–2), (2) number of peeks, and (3) latency to first peek. Scores were standardized (with number of peeks reverse scored) and summed to create a composite gift delay score. In contrast to the other measures, children's performance on gift delay was not significantly correlated with age, $r(44) = .17, ns$.

Relations among executive functioning tasks. Performance on the three executive function measures was significantly intercorrelated. The whisper measure was correlated with both the bear/dragon measure, $r(41) = .50, p < .05$, and the gift delay measure, $r(44) = .49, p < .05$. Also, performance on the bear/dragon task was correlated with performance on the gift delay task, $r(41) = .31, p < .05$. These correlations remained significant when age was controlled (all $r_s > .28, p_s < .05$).

Relations Between Executive Function, False Photographs, and False Beliefs

The raw and age-controlled correlations between the false belief, false photograph, and executive functioning tasks are presented in Table 2. False belief performance was correlated with performance on bear/dragon in both raw and age-controlled

correlations. In addition, false belief was marginally related to gift delay, although this relation was smaller when age was controlled. In contrast, performance on false photograph was not significantly correlated with any of the executive functioning variables. Finally, the magnitude of the correlation between false belief and bear/dragon remained strong when false photograph performance was controlled, $r(38) = 0.45, p < .01$.

Discussion

The main purpose of Study 1 was to assess whether children's abilities to reason about both false photographs and false beliefs are related to performance on executive functioning tasks. Results showed that only reasoning about false beliefs was significantly related to children's executive functioning skills; reasoning about false photographs was not. Consistent with previous research (Carlson & Moses, 2001; Carlson et al., 2002, 2004; Hala et al., 2003), the relation between executive functioning and false belief was strongest for the bear/dragon task that requires choosing among conflicting response options. This pattern was present whether or not age was held constant.

In contrast to some other studies (e.g., Leslie & Thaiss, 1992; Zaitchik, 1990), children found false belief somewhat more difficult than false photograph (see also Slaughter, 1998). This raises the potential concern that our ability to detect correlations between executive functioning and false photograph may have been constrained by ceiling effects or a lack of variability on false photograph. However, this seems unlikely. First, children were clearly well below ceiling performance on false photograph. Second, the standard deviations for false belief and false photograph tasks are roughly equivalent. These findings give us confidence that correlations between executive functioning and false photograph reasoning could have emerged had they existed.

Table 2
Raw/Age-Controlled Correlations (With Valid N) Between False Belief, False Photograph, and Executive Functioning Tasks

Representation task	Executive function task		
	Bear/dragon (aggregate)	Gift delay (aggregate)	Whisper
False belief	.48* / .30* (41)	.26 [†] / .22 (44)	.22 / .16 (44)
False photograph	.19 / .03 (41)	.12 / .07 (43)	.05 / – .02 (43)

Note: * $p < .05$, [†] $p < .10$.

Nonetheless, an important goal of Study 2 was to replicate the differential relations of executive functioning skills and children's emerging abilities to reason about false beliefs and false photographs. If the pattern of results holds, a central question concerns why children's executive skills relate to their abilities to reason about false beliefs but not false photographs. As noted earlier, there are at least two possible explanations for this dissociation. The first is that beliefs are arguably more abstract than photographs, which have clear physical instantiations. This difference in abstraction might lead to differential demands on executive functioning skills. For instance, reasoning about beliefs may impose a heavier demand on working memory skills because there is no physical instantiation of the representation (like a photograph) that could act as a cue to retrieval. Similarly, the physical instantiation of the photograph may make the photograph more similar to reality in terms of salience than is the case for beliefs. Hence, reasoning about mental states might impose a heavier requirement to distance oneself from a salient reality in order to reflect on the mental representation itself (Russell et al., 1999). Negotiating either of these additional demands could potentially tax children's emerging executive functioning skills.

The second way in which beliefs and photographs differ concerns the function of beliefs versus photographs as representations of the world. That is, the individual who holds a belief intends that belief to be a faithful up-to-date representation of some true state of affairs. As we learn more about the world, we adjust our beliefs accordingly. Photographs, in contrast, are not usually expected to be up-to-date representations of the world after they have been taken. Instead, they are intended to capture faithfully the state of the world at a specific point in time. Subsequent changes to the world have no bearing on the content of photographic representations. This difference in how mental and photographic representations relate to the world may account for the increased executive demands present for false belief reasoning; in order to think about a false belief, children may have to override a prepotent assumption regarding beliefs being up-to-date representations of the world.

These two general hypotheses regarding the relation between executive functioning and false belief offer specific predictions about the extent to which executive functioning might be important for reasoning about any kind of representation. If executive function is required to think about representations that are highly abstract, we might expect that executive skills would only be implicated in reasoning

about representations that do not have a clear physical instantiation (e.g., mental states). Alternatively, it is possible that executive skills are required to reason about representations that are intended to be truthful up-to-date representations of the world, irrespective of whether or not a physical instantiation is present. The goal of Study 2 was to test these hypotheses by examining the relation between executive functioning and children's understanding of "false signs."

Study 2

Nonmental representational media vary in the extent to which they are meant to represent faithfully a current real situation. As we have argued, it is not surprising when a previously taken photograph no longer accurately represents a current state of affairs. In contrast, however, directional signs—such as an arrow pointing to the washrooms in a restaurant—are intended to maintain correspondence with reality. That is, the "job descriptions" of a belief and a sign are similar in that they are both meant to represent truthfully a current state of affairs. Parkin and Perner (1996) developed a task for testing preschoolers' understanding of false signs that closely paralleled the structure of a false belief task in that a sign became outdated as the real state of affairs changed. Intriguingly, there were robust correlations between preschoolers' performance on the false sign and false belief measures.

Importantly, for the purposes of the present study, the false sign task provides an elegant way of further clarifying the nature of the relation between false belief and executive functioning. On the one hand, signs are like photographs and unlike beliefs in that they have a clear physical instantiation. Because of their physical instantiation, we might expect that false sign reasoning would impose relatively minimal working memory demands and minimal requirement to distance from reality to reflect on the representation, thereby resulting in a nonsignificant correlation with executive functioning. On the other hand, if executive functioning is important because it enables children to consider that a usually true representation might be false, we would expect a robust correlation between executive functioning and false sign reasoning, despite the sign's physical instantiation. Like beliefs, but unlike photographs, signs are meant to be faithful up-to-date representations of some real-world situation.

Thus, there were two main goals of Study 2. The first was to retest the relation between false beliefs, false photographs, and executive functioning to

determine whether the same general pattern of findings as found in Study 1 would emerge. The second was to include the false sign task to gain insight into the nature of the relation between executive functioning and false belief. In addition to these two goals, we added a measure of children's verbal ability in order to more precisely specify the relations between executive functioning and children's understanding of different kinds of false representations.

Method

Participants

Sixty preschool children participated initially but 6 were omitted because they failed to complete the entire battery of tasks. Thus, the final sample consisted of 54 preschool children (25 female) aged between 3 years 1 month and 4 years 10 months ($M = 3$ years 11 months). For certain analyses, participants were divided into two age groups based on median split. For simplicity, we will refer to these as the younger ($M = 3$ years 4 months; range: 3 years 1 month to 3 years 8 months) and older ($M = 4$ years 5 months; range: 3 years 9 months to 4 years 10 months) children. The sample was recruited from a database drawn from a primarily White middle-class university community in southeastern Ontario, Canada. These children were tested in an on-campus playroom.

Measures

This study included the false belief, false photograph, and executive functioning measures from Study 1. Interrater reliability for each of the three executive measures was calculated on 20% of the transcripts, and in each case was found to be acceptable (bear/dragon: 99.2% agreement, $\kappa = .97$; whisper: 82.9% agreement, $\kappa = .73$; gift delay-resistance: 75% agreement, $\kappa = .60$; gift delay-peeks: 83% agreement, $\kappa = .76$; gift delay-latency: 83% within 2 s). Internal reliability of the executive function measures was also high (bear/dragon: $\alpha = .92$; whisper: $\alpha = .88$; gift delay: intercorrelations = .88–.92). In addition, we added two false sign tasks and the Peabody Picture Vocabulary Test (PPVT–R) (Dunn & Dunn, 1981). The tasks were administered by one of two female experimenters. Methods of administration and scoring were identical to those outlined earlier. The new measures are described in detail below.

False sign: location change. This task, adapted from Parkin and Perner (1996), was modeled after the corresponding false belief and false photograph tasks. In a brief training period, the experimenter introduced children to an arrow mounted on a cardboard card (2.5×7 cm) that pointed toward either a blue or a red cardboard house. The experimenter explained that a story character (Chester, a figurine) used the arrow to let everyone know where he was playing. The experimenter then ensured that children understood by testing them four times (i.e., "Where does the arrow say Chester is?"). All children clearly demonstrated that they understood the arrow sign.

Children then heard a story introducing Chester and his friend Marianne (another figurine) who were trying to decide whether they wanted to play in the red or blue house. Then, Marianne got hungry and wanted to get a snack, to which Chester replied, "That's okay, I'll leave the arrow pointing to where you can find me." Then, when Marianne went away Chester pointed the arrow at the red house, went inside, and was no longer visible. The experimenter explained that after awhile, Chester decided that he'd rather be in the blue house, and then changed locations. But, "Silly Chester. He forgot to change the sign!" Marianne then returned, wanting to find Chester. Children were then asked two questions: (1) *False Sign Test Question*: "Where does the sign say Chester is?" and (2) *Reality Control Question*: "Where is Chester really?" The arrow remained in full view while these questions were asked. Children responded by pointing to one of the two houses. Children who failed the control question on the false sign tasks were omitted from analyses. Children who answered both the control and test question correctly received a score of 1.

False sign: contents change. In this task, participants were introduced to a puppet (Bert) who was portrayed putting his pet cat (a small figurine) inside a box. He then put a sign on the box showing a picture of a cat, so that he could remember what was inside. Then Bert went away. While Bert was gone, the cat jumped out of the box and a frog jumped in. Bert then returned and children were asked two questions: (1) *False Sign Test Question*: "What does the sign say is in the box? A cat or a frog?," and (2) *Reality Control Question*: "What is really inside the box? A cat or a frog?" Children who passed both questions received a score of 1.

PPVT–R. This is a standardized test of children's receptive vocabulary development, and is generally taken as an excellent indicator of children's verbal mental age. In the test trials, children are asked to

Table 3
Means and Standard Deviations for Tasks (With Range of Scores) Used in Study 2

		Executive functioning tasks																	
		Representation tasks						Bear/dragon				Gift delay						Whisper (0–2)	
		False belief (0–2)		False photo (0–2)		False signs (0–2)		Practice (1–5)		Total (0–15)		Peek resistance (0–2)		Total # of peeks (0–9)		Latency (0–60 s)			
Age group	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Younger	26	0.46	.65	0.92	.63	0.50	.65	.35	.49	6.65	5.69	0.84	.88	2.92	2.43	26.69	25.20	1.36	.77
Older	28	1.21	.83	1.39	.73	1.50	.63	.78	.42	11.89	5.24	1.21	.92	1.39	2.01	45.61	21.58	1.66	.59

point to the picture corresponding with a given word. The target picture is located in a 2×2 grid in which the other three cells contain distracters. Children begin the test at an age-appropriate level and then proceed through progressively more difficult sets of 10 pictures each. The test is discontinued after children fail six items within a set. Raw scores are assigned on the basis of how many blocks children complete, adjusting for errors they may have made along the way.

Design

The tasks were administered in one session that lasted approximately 45 min. As in Study 1, the tasks were administered in a fixed order: (1) PPVT, (2) false sign: location, (3) false belief: contents, (4) bear/dragon, (5) false photograph: contents, (6) gift delay, (7) false belief: location, (8) whisper, (9) false photograph: location, and (10) false sign: contents. The sessions were videotaped and children's responses were scored from videotape.

Results

The two main goals of Study 2 were to (1) replicate the principal findings from Study 1 and (2) add the false sign task to the battery of tasks to determine whether executive functioning is required to reason about representations that are abstract (i.e., no physical instantiation) or for any representation that is supposed to update in order to represent reality faithfully.

Comparison of False Belief, False Photograph, and False Signs

Preliminary analyses revealed that performance across the two trials within each task did not differ.

Thus, scores on the two trials were combined to create a single measure for each of the tasks (range for each: 0–2). The means and standard deviations for these measures are presented in Table 3.

A 2 (age) \times 3 (task) mixed-design ANOVA with task as a within-subjects factor revealed a significant main effect of age, $F(1, 52) = 56.25$, $p < .05$, $\eta^2 = .52$, and a significant main effect of task, $F(2, 104) = 3.10$, $p < .05$, $\eta^2 = .06$. However, the significant main effect of task was qualified by a significant Age \times Task interaction, $F(2, 104) = 3.51$, $p < .05$, $\eta^2 = .06$. Follow-up paired t tests showed that younger children found false photograph easier than both false belief, $t(29) = 3.57$, $p < .05$, and false sign, $t(29) = 2.18$, $p < .05$. There were no significant differences in performance across the tasks for the older children. Moreover, the age effects were significant within each of the tasks, $t_s(52) > 2.51$, $p_s < .02$.

False belief was correlated with both false photograph, $r(52) = .39$, $p < .05$, and false sign, $r(52) = .53$, $p < .05$. False photograph was moderately correlated with false sign, although this relation only reached marginal significance levels, $r(52) = .25$, $p = .07$. When these analyses were controlled to account for general improvements owing to age, the relation between false belief and false sign fell to a marginal significance level, $r(50) = .24$, $p = .08$, and all others were nonsignificant, $p > .10$. When both age and PPVT were controlled, there were no significant relations among the false representation tasks.

Executive Functioning Measures

The means and standard deviations for each of the executive functioning tasks are presented in Table 3. As in Study 1, preliminary analyses showed that all of the dependent measures within the bear/dragon and gift delay tasks were intercorrelated ($r_s .48-.92$). Thus, these dependent measures were standardized

Table 4

Raw/Age-Controlled/Age- and PPVT-Controlled Correlations (With Valid N) Between Representation and Executive Function Tasks

Representation task	Executive function task		
	Bear/dragon (aggregate)	Gift delay (aggregate)	Whisper
False belief	.51**/.28*/.26 [†] (51)	.30*/.13/.10 (54)	.28*/.11/.11 (53)
False photograph	.25 [†] /.06/.05 (51)	.13/-.06/-.08 (54)	.16/.06/.06 (53)
False sign	.58**/.33*/.32* (51)	.27*/.04/.00 (54)	.14/-.07/-.08 (53)

Note: ** $p < .01$, * $p < .05$, [†] $p < .10$, all significance levels two-tailed.

and summed to create a composite score for each task. As in Study 1, age was correlated with performance on both the bear/dragon and whisper tasks, $r(49) = .58$, $p < .05$, and $r(51) = .32$, $p < .05$, respectively. In this study, age also correlated with performance on whisper, $r(52) = .34$, $p < .05$. Children's PPVT scores were also correlated with performance on bear/dragon, $r(49) = .43$, $p < .05$, and gift delay, $r(52) = .27$, $p < .05$, but were only marginally correlated with whisper, $r(51) = .23$, $p < .10$. Interestingly, in contrast to Study 1, we did not find significant intercorrelations among the executive functioning tasks.

Relations Between False Belief, False Photograph, False Sign, and Executive Functioning

The raw, age-controlled, and age plus PPVT-controlled correlations between the false belief, false photograph, false sign, and executive functioning measures are summarized in Table 4. The key findings were that performance on both false belief and false signs was correlated with performance on the bear/dragon task, but performance on false photograph was not. This pattern was significant after controlling for age, and controlling for age and PPVT simultaneously, although in the latter analysis, the correlation involving false belief fell to marginal significance levels ($p = .07$). There was some evidence to suggest that performance on false belief and false sign also correlated with gift delay, and that false belief correlated with whisper. However, these relations did not survive the more stringent analyses. Finally, performance on bear/dragon predicted both false belief and false sign performance while simultaneously controlling for age, PPVT, and performance on false photograph [false belief—bear/dragon: $r(46) = .26$, $p < .05$, one-tailed; false sign—bear/dragon: $r(46) = .32$, $p < .05$, one-tailed].

Analyses of False Photograph and False Belief in Combined Studies

The findings from Studies 1 and 2 were consistent in that they both revealed a strong relation between children's bear/dragon performance and reasoning about false beliefs, but not about false photographs. An important question concerns whether correlations between false photograph performance and executive functioning might also be present but relatively weak, and thus difficult to detect because of insufficient power. To gain evidence regarding this possibility, we combined the samples from Studies 1 and 2 to conduct a more powerful test of the relation between the executive functioning battery, false belief, and false photograph.

Preliminary analyses showed that children's age-in-months was a significant predictor of children's performance on the false belief ($r = .61$, $p < .01$), false photograph ($r = .39$, $p < .01$), bear/dragon ($r = .50$, $p < .01$), and gift delay ($r = .25$, $p < .05$) tasks. Age was a near-significant predictor of performance on the whisper task ($r = .18$, $p = .07$). The raw and age-controlled correlations between performance on the executive functioning measures, false belief, and false photograph are summarized in Table 5. We also added to this analysis an aggregate measure combining all three of the executive measures. These more powerful analyses revealed a significant raw correlation between performance on bear/dragon and false photograph. However, once age was statistically controlled, this relation plummeted to nearly 0. As expected, the correlation between bear/dragon and false belief was robust in this more powerful analysis, and the same was true for the executive functioning aggregate. Moreover, there were suggestions that the other measures of executive functioning also related to false belief, although not as strongly as bear/dragon.

Table 5

Raw/Age-Controlled Correlations (with Valid N) for False Photograph and False Belief with Executive Functioning Tasks in Combined Studies

Representation task	Executive function task			
	Bear/dragon (aggregate)	Gift delay (aggregate)	Whisper	EF aggregate
False belief	.50**/.26* (92)	.28*/.18 ^t (98)	.24*/.17 ^t (98)	.48**/.29** (92)
False photograph	.23*/.05 (92)	.13/.04 (97)	.06/.01 (97)	.19 ^t /.04 (92)

Note:

** $p < .01$, * $p < .05$, ^t $p < .10$, all significance levels two-tailed.

Correction for attenuation. A critical question concerns whether differences in the pattern of correlations might be accounted for by differences in the psychometric reliability of the measures that were used. Most importantly, it is possible that false belief performance correlated more highly with the executive measures because it is simply a more reliable measure than the false photograph task. If this were true, the expectation is that if the measures had been of similar reliability, they would show similar correlations with executive functioning measures. To gain insight into whether this is a plausible explanation of our findings, we computed the “true correlation” for all of our relevant variables using the formula for “correction for attenuation” (Lord & Novick, 1968; Nunnally, 1967). The correction for attenuation (see Appendix A) is a measure widely accepted by psychometricians for estimating the true correlation between two measures by correcting for attenuation in the observed correlation caused by measurement error.

In order to compute the true correlations, we needed to estimate the reliability for the two measures. For the executive functioning tasks, we used the average reliability of the measures across the two studies. For false belief and false photograph, we did not have sufficient trials to compute standard reliability measures, nor did we have a measure of test–retest reliability. However, the false belief and false photograph task each showed significant correlations with age, thereby demonstrating their external validity. Because variables can only be valid insofar as they are reliable, the correlations with age serve as *lower-bound* estimates of the reliability of our false belief and false photograph measures. Using these reliability estimates, we computed the true age-controlled correlations for all the relevant variables in the combined studies analysis. These data are summarized in Table 6. Note that because we used a lower-bound estimate of our measures’ reliabilities, the true correlations are *upper-bound* estimates (i.e.,

maximally corrected) of the actual true correlation between measures.

Two points are noteworthy about these analyses. First, for all executive functioning measures, the estimated true correlations with false belief are substantially higher than the correlations for false photograph. Second, even with these upper-bound estimates of the true correlations, correlations involving the false photograph measure remain very low.

We also conducted a Monte Carlo simulation (details of which are available from the first author) showing that the correlations between false photograph and the executive functioning measures would be substantially higher if their attenuation, relative to the correlations involving false belief, were due only to their poorer reliability. Together, these findings give us confidence that potential differences in measure reliability alone cannot account for the lack of correlation between false photograph and executive functioning.

Summary

The main goals of Study 2 were to (1) replicate the finding that executive functioning is related to performance on false belief but not false photograph

Table 6

Age-Controlled Correlations (With Valid N) for False Photograph and False Belief with Executive Functioning Tasks in Combined Studies With Correction for Attenuation

Representation task	Executive function task			
	Bear/dragon	Gift delay	Whisper	EF aggregate
False belief	.36 (92)	.25 (98)	.24 (98)	.39 (92)
False photograph	.09 (92)	.07 (97)	.02 (97)	.09 (92)

and (2) gain insight into why executive functioning might be important for reasoning about false beliefs. To this end, we replicated our first study while adding false sign to the battery. The results showed that children's performance on executive functioning tasks, in particular bear/dragon, was correlated with false sign and false belief, but not with false photograph. When taken together, these findings support the hypothesis that executive functioning is related to reasoning about representations that are supposed to be up-to-date representations of some true state of affairs, regardless of whether the nature of the representation is concrete (like a sign) or abstract (like a belief).

General Discussion

Executive Functioning and Reasoning About Representations

Previous research has clearly demonstrated ontogenetic and neurocognitive dissociations between false belief and false photograph reasoning (Leslie & Thaiss, 1992; Sabbagh & Taylor, 2000). The first goal of the present research was to explore the source of these dissociations by examining whether false belief and false photograph relate differentially to executive functioning. On this point, we found that performance on bear/dragon—the executive functioning task that is typically most robustly associated with theory-of-mind development (Carlson & Moses, 2001; Carlson et al., 2002; Hala et al., 2003)—correlated with performance on false belief but not false photograph tasks. This general pattern of performance was replicated across two studies, remained when age and verbal ability were held constant in Study 2, and was also true when the samples from both studies were combined in a single, more powerful analysis. This finding suggests that false belief and false photograph tasks differ in the extent to which they each tap the executive ability measured by the bear/dragon task—namely the ability to selectively inhibit responding according to a rule.

We reasoned that judgments about false beliefs might pose greater executive demands relative to false photographs for two reasons: (1) beliefs are more abstract than photographs and (2) beliefs are supposed to represent reality faithfully, even after time has passed, while photographs are not. To test between these two possibilities, we designed Study 2 to replicate the findings from Study 1 and explore the links between executive functioning and children's understanding of false signs. False signs represent an

important test case because like photographs they are concrete, but like beliefs they are supposed to represent faithfully the state of the world. We found that individual differences in the bear/dragon task significantly predicted performance on false sign tasks, thereby supporting the hypothesis that aspects of executive functioning may be required to think about any kind of representations (mental or non-mental) that are supposed to be up-to-date with respect to reality.

Conversely, our findings speak against the hypothesis that executive functioning is only required to reason about representations that are abstract because of the additional demands that abstract representations might impose (e.g., distancing from reality, working memory). In the false sign task, the sign has a physical instantiation that is even more salient than that of the photograph because its content (e.g., the direction of the arrow) was immediately available to children. Yet, performance on false sign was related to executive functioning whereas false photograph performance was not. This strengthens the case against the possibility that mental representations are difficult because of their abstractness. Further evidence in this regard comes from recent findings showing that children's executive skills may be specifically important for reasoning about beliefs and not for reasoning about other kinds of mental representations, such as desires (Moses, Carlson & Sabbagh, 2005). Whereas beliefs and desires are both abstract mental states, they differ in terms of whether they are generally thought to be an up-to-date, truthful representation of reality—beliefs are meant to be true whereas desires are not (Moses, 1993; Perner, 1991a).

Our proposal also contrasts with accounts that suggest that the relation between executive functioning and false belief can be attributed to task complexity. For instance, Zelazo and colleagues (Frye et al., 1995; Zelazo, 2000) have argued that the false belief task requires children to reason in terms of hierarchically embedded rules, which in turn poses heavy demands on executive functioning. However, in terms of formal rule structure the false belief and false photograph tasks would appear to be equivalent (although see Muller, Zelazo, & Imrisek, 2005). On the basis of our analysis, we suggest that where the two tasks differ is in the relation between the representation and the state of the world.

Limitations

Some potential conceptual and methodological caveats need to be considered in evaluating our

general proposal. First, our claim that photographs impose less demand on executive function is predicated on the assumption that children recognize that photographs are not typically updated to reflect current reality. Clearly, general observation comports with this assumption—that is, preschoolers do not seem in any way bewildered by photographs that depict a scenario other than what is currently true. Nonetheless, demonstrating this understanding empirically would lend further support to our explanation for why false beliefs and false signs require executive functioning, but false photographs do not.

A second concern is whether reasoning about false signs is a truly “nonmental” task. Some theorists have suggested that signs, such as a directional arrow, have meaning by virtue of their having been designed to fulfill some communicative intention (e.g., Searle, 1983). Perhaps, then, theory-of-mind skills are necessary to solve the false sign task because children have to reason about the thwarted communicative intention of the character who designed the sign. While we cannot rule this possibility out entirely, we do not find it plausible. For a start, not all theorists agree that understanding signs requires a concomitant understanding of intentions, particularly when the signs at stake are highly transparent, such as directional arrows (e.g., Wittgenstein, 1953/1997). In this regard, it is worth noting that “natural” signs, such as smoke as an indicator of fire, do not rest on any assumptions about communicative intent. Hence, there is no reason to suppose that mental state understanding is required to make sense of false signs in general.

Third, from a methodological perspective, it is important to note that our key conclusions are based, at least in part, upon the absence of a relation between false photograph and our executive functioning measures. As with any null effect, there are several possible artifactual interpretations. As noted previously, one possibility is that the false photograph task may be somewhat less reliable (or “noisier”) than the false belief or false sign, thereby making it less amenable to detecting relevant individual differences. We cannot entirely rule out this possibility because no studies have compared the test–retest reliability of these representation tasks. However, our analyses in which we re-computed the correlations using the correction for attenuation procedure showed that substantial differences remain between the false belief relative to the false photograph correlations even when statistically correcting for potential differences in reliability. Nonetheless, establishing test–retest reliability for these measures would be a further important step in establishing further confidence in this pattern of findings.

Implications for Autism

These caveats notwithstanding, our findings have important implications for understanding the nature of the theory-of-mind impairment in autism. The dissociation between false belief and false photograph tasks has been particularly important in supporting the notion that individuals with autism have a “specific” deficit in theory-of-mind reasoning. For example, Leslie and Thaiss (1992) have argued that through generally poor performance on a false belief task and generally strong performance on a false photograph task, individuals with autism reveal a specific difficulty with reasoning about others’ mental states independent of any domain-general cognitive impairments. However, the present findings raise the possibility that some of this deficit may be attributable to the fact that, because of their conceptual characteristics, false belief tasks make executive demands that are not made by false photograph tasks. A number of researchers have now shown that autistic individuals are impaired on a variety of executive tasks (Russell, 1997). Thus, the present findings leave open that the domain-specific false belief impairment in autism may actually be related to a more domain-general executive functioning deficit.

On a more constructive note, our findings suggest that a more appropriate test of a strong version of the “domain-specificity” hypothesis may come from a comparison of performance on false belief and false sign tasks. Given that false beliefs and false signs share conceptual characteristics that make them especially reliant on executive functioning, any dissociation in performance seen in a neuroimaging study, or in a study with autistic individuals, might be more confidently attributed to the fact that reasoning about false beliefs requires theory of mind.

Finally, a similar point can be raised with respect to understanding the neural bases of theory of mind. A number of recent reviews have suggested that left medial prefrontal regions may make a particularly important contribution to theory-of-mind reasoning (Frith & Frith, 1999). In an event-related potential (ERP) study, this general pattern was found when adults’ reasoning about false beliefs and about false photographs were explicitly compared (Sabbagh & Taylor, 2000). In light of the present findings, an important question concerns the extent to which these domain-specific dissociations may instead reflect differences in the extent to which different kinds of representations tap domain-general processes. Considerable neuropsychological evidence suggests that prefrontal regions are critical for executive

functioning in general, and inhibitory control in particular (Diamond, 2002; Luria, 1973; Miller, 2000). Because the extent to which these regions overlap with those critical for theory-of-mind understanding has not been fully investigated, the present findings militate for future research in this intriguing area.

Conclusion

We found evidence that children's reasoning about false beliefs and false signs is predicted by their emerging executive functioning skills, whereas their reasoning about false photographs is not. This pattern of findings suggests that executive functioning skills are heavily recruited any time children reason about representations that are supposed to remain faithful to the changing state of the world. These findings have implications for understanding both the domain-general and domain-specific developments that are critical for theory-of-mind development in the preschool years.

References

- Bartsch, K., & Wellman, H. M. (1994). *Children talk about the mind*. Cambridge, UK: Cambridge University Press.
- 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.
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive function and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299–329.
- 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.
- Carpendale, J. I. M., & Lewis, C. (2005). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79–151.
- Charman, T., & Baron-Cohen, S. (1992). Understanding drawings and beliefs: A further test of the metarepresentation theory of autism: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 1105–1112.
- Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25–31.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function*. New York: Oxford University Press.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised: Form M*. Circle Pines, MN: American Guidance Service.
- Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science, 286*, 1692–1695.
- 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.
- 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.
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.
- 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.
- Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40*, 203–218.
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*, 59–65.
- Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development, 8*, 1–25.
- Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2005). On the specificity of the relation between executive function and children's theory of mind. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelations among executive functioning, working memory, verbal ability and theory of mind* (pp. 131–145). Mahwah, NJ: Lawrence Erlbaum Associates.
- Muller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive functioning and children's understanding of false belief: How specific is the relation. *Cognitive Development, 20*, 173–189.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

- Parkin, L. J., & Perner, J. (1996). *Wrong directions in children's theory of mind: What it means to understand belief as representation*. Unpublished manuscript, University of Sussex.
- Perner, J. (1991a). On representing that: The asymmetry between belief and desire in children's theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind* (pp. 139–155). Hillsdale, NJ: Lawrence Erlbaum.
- Perner, J. (1991b). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3, 337–344.
- Reed, M., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly*, 30, 131–147.
- Repacholi, B. M., & Slaughter, V. (2003). *Individual differences in theory of mind: Implications for typical and atypical development*. Hove, UK: Psychology Press.
- Russell, J. (1996). *Agency*. Cambridge: Lawrence Erlbaum Associates/Taylor & Francis Ltd.
- Russell, J. (Ed.). (1997). *Autism as an executive disorder*. Oxford, UK: Oxford University Press.
- Russell, J., Saltmarsh, R., & Hill, E. (1999). What do executive factors contribute to the failure on false belief tasks by children with autism? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40, 859–868.
- Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science*, 11, 46–50.
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). Executive functioning and theory-of-mind in preschool children from Beijing, China: Comparisons with U.S. preschoolers. *Psychological Science*, 17, 74–81.
- Searle, J. R. (1983). *Intentionality*. Cambridge, UK: Cambridge University Press.
- Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, 69, 321–332.
- Taylor, M. (1996). A theory of mind perspective on social cognitive development. In R. Gelman & T. Au (Eds.), *Perceptual and cognitive development. Handbook of perception and cognition* (2nd ed., pp. 283–329). San Diego, CA: Academic 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.
- 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.
- Wittgenstein, L. (1953/1997). *Philosophical investigations*. Oxford, UK: Blackwell.
- Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35, 41–68.
- Zelazo, P. D. (2000). Self-reflection and the development of consciously controlled processing. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 169–189). Hove, UK: Psychology Press.

Appendix A

Correction for attenuation

When a given measure exhibits less-than-perfect internal reliability, its "true" predictive validity is attenuated by the square root of its reliability (Lord & Novick, 1968; Nunnally, 1967). Thus, the true correlation between any two measures can be computed from the following equation:

$$r_{tx,ty} = \frac{r_{ox,oy}}{\sqrt{r_{xx}r_{yy}}}$$

where $r_{tx,ty}$ is the true correlation between x and y , $r_{ox,oy}$ is the observed correlation between x and y , r_{xx} is the reliability of the x scores, and r_{yy} is the reliability of the y scores.

When reliabilities for x and y are both 1, the observed and true correlations between x and y are equal. More important, as reliabilities of x or y decrease, the observed correlation between x and y is attenuated relative to its true correlation. The correction for attenuation thus provides a way of comparing the true predictive validity of measures with differences in internal reliability.