



Original Articles

Executive function depletion in children and its impact on theory of mind

Lindsey J. Powell ^{a,b,*}, Susan Carey ^b^a Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States^b Department of Psychology, Harvard University, Cambridge, MA, United States

ARTICLE INFO

Article history:

Received 30 October 2012

Revised 5 January 2017

Accepted 29 March 2017

Available online 18 April 2017

Keywords:

Executive function

Theory of mind

Cognitive development

Ego depletion

ABSTRACT

The current studies provide an experimental, rather than correlational, method for testing hypotheses about the role of executive function (EF) in conceptual development. Previous research has established that adults' tendency to deploy EF can be temporarily diminished by use. Exercising self-control in one context decreases adults' performance on other EF demanding tasks immediately thereafter. Using two different depletion methods, Experiments 1 and 3 extend this finding to preschool-aged children. Experiments 2 and 4 make use of these EF depletion methods to elucidate the role of EF in children's theory of mind reasoning. Experiment 2 shows that EF depletion affects 5-year-olds' ability to predict another's behavior on the basis of that person's false belief, and Experiment 4 shows that this negative effect of depletion extends to 4- and 5-year-olds' ability to explain others' behavior on the basis of their false beliefs. These findings provide direct evidence that EF is required for the expression of an understanding of others' false beliefs across a variety of task demands, even in children who clearly have the capacity to construct such representations. We suggest ways in which depletion may be used as a tool for further investigating the role of executive function in cognitive development.

© 2017 Published by Elsevier B.V.

1. Introduction

The hypothesis of a 'central executive' or a set of executive functions (EF) was introduced by neuropsychologists as they sought to explain the damage done by lesions to the frontal lobe, which often result in subtle but devastating effects on the ability to plan and make everyday decisions (Shallice & Burgess, 1991). Research on healthy adults has helped psychologists to dissect EF into partially separable component processes, including inhibition, working memory, and task- or set-switching abilities. These processes often operate together to allow for the execution of complex cognitive processes and behavior (Miyake et al., 2000).

1.1. EF and cognitive development

Recently, developmental research has begun to show just how crucial EF resources are for learning. Measures of EF correlate with teachers' assessments of 'school readiness' and with students' academic performance (Blair & Razza, 2007). Moreover, EF skill corre-

lates with children's performance on tests of understanding in both academic and non-academic domains, including theory of mind, math, biology, and physical reasoning (Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011; Bull & Scerif, 2001; Carlson & Moses, 2001; Zaitchik, Iqbal, & Carey, 2013). These correlations persist even when age and verbal intelligence are controlled for, suggesting that EF may have a direct relationship with knowledge acquisition and use.

Such findings have spurred psychologists and educators to begin to design EF training programs for classrooms (e.g. Diamond, Barnett, Thomas, & Munro, 2007), but as they do so it would be useful to have a clearer picture of *how* EF relates to the acquisition and use of new knowledge. Though the correlational research referenced above is persuasive regarding the existence of a relationship, it cannot tell us what role EF plays in learning or even, in some cases, what the direction of causation is. Even under the assumption that the maturation of EF plays a role in driving conceptual development, a correlation between EF and performance in any particular domain is compatible with a role for EF in either the *construction* of a particular body of knowledge or the *selective application and expression* of that knowledge once it has been acquired, or both. Unfortunately, we currently lack experimental methods we can use to directly test the role of EF in children's learning and reasoning processes.

* Corresponding author at: Department of Brain and Cognitive Sciences, Room 46-4021, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States.

E-mail address: ljpowell@mit.edu (L.J. Powell).

The best test of an expression account of the correlations between measures of EF and measures of conceptual understanding would be to experimentally manipulate the EF of a group of participants who normally show evidence of having the knowledge in question. Evidence that participants randomly assigned to a low EF condition perform worse on relevant tasks than participants in a high EF condition would show that, even after the acquisition of the knowledge in question, EF capacity affects its use. It is likely that the reason this approach has not been taken in the past is that developmental researchers have viewed EF as a stable trait or skill that, while trainable over long periods of time, is not malleable within the scope of a single experimental session. This assumption turns out to be false. Recent research with adults has shown that EF can be temporarily depleted with use. Participants who complete a task involving heavy EF demands do worse on a subsequent EF-laden task than participants who begin with an easy task that places minimal demands on EF (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Schmeichel, 2007; for reviews see Hagger, Wood, Stiff, & Chatzisarantis, 2010; Hofmann, Schmeichel, & Baddeley, 2012). Here we adapt this experimental paradigm for use with children and then to use it to test whether EF is needed for preschoolers' expression of their theory of mind.

1.2. EF and theory of mind

Theory of mind refers to the lens through which human adults view one another, explaining behavior by appealing to mental states like thoughts, feelings, and goals. Many of the social cognitive capacities that comprise a full theory of mind begin to emerge in infancy, but one central component – an explicit understanding of beliefs – appears much later, around 3 or 4 years of age (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). This is a striking delay, and the fact that preschoolers are also undergoing substantial improvements in multiple areas of EF has not gone unnoticed. A number of studies have demonstrated a strong correlation between young children's ability to reason about beliefs and their EF skills (e.g. Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Hughes, 1998). This research has inspired both expression and construction accounts of how EF maturation may lead to more successful belief reasoning, as well as additional theories questioning whether causation may run in the opposite direction (e.g. Perner & Lang, 1999) or be related to a third, unmeasured variable such as hierarchical reasoning abilities (e.g., Frye, Zelazo, & Palfai, 1995) or the maturation of dopaminergic systems in the frontal lobes, which in turn contributes to the maturation of both EF and theory of mind, independently (Lackner, Bowman, & Sabbagh, 2010).

The expression hypothesis is partly motivated by an analysis of the task demands associated with preschool measures of theory of mind. Clear evidence of belief understanding often involves reasoning about beliefs that conflict with reality, because it is in these cases that belief-based and reality-based predictions diverge (Dennett, 1978). For example, false belief tasks feature a protagonist who is mistaken about some fact, such as the location of a toy, and require the participant to predict the protagonist's thoughts or actions on the basis of this false belief. This methodological constraint means that passing tests of belief understanding requires more than just a functioning concept of beliefs (Bloom & German, 2000). Even assuming they represent the protagonist's belief, children must maintain both this representation and that of the actual location of the toy to follow the story and may have to inhibit the latter representation in order to base a judgment on the former. Moreover, both superficial aspects of the task, such as the need to point to an empty location when a salient object is nearby, and intrinsic aspects, such as the need to select between candidate representations of another's beliefs,

may place further demands on inhibitory control (Carlson, Moses, & Hix, 1998; Leslie & Polizzi, 1998). Given such demands, it seems likely that canonical theory of mind tasks draw directly on main components of EF, including working memory and inhibitory control, and young preschoolers may simply lack the relevant EF to succeed. Indeed, when EF demands are increased, older children and even adults become more likely to fail tests of belief understanding (German & Hehman, 2006; Leslie, German, & Polizzi, 2005).

In fact, some researchers argue that the EF demands of preschool theory of mind tasks are the *only* thing masking an understanding of beliefs, even false beliefs, that is present from infancy (e.g. Leslie, 1994). Many recent studies show that even infants implicitly predict the actions of other agents on the basis of the information available to those agents, rather than on the basis of current reality (e.g. Onishi & Baillargeon, 2005; Song, Onishi, Baillargeon, & Fisher, 2008; Surian, Caldi, & Sperber, 2007; for a review see Baillargeon, Scott, & He, 2010). To the extent that these findings reflect a rich understanding of beliefs present from the second year of life on, there is no need for preschoolers to construct a new understanding of beliefs and thus no construction process for EF to play a role in. Researchers holding this point of view conclude that the correlations between EF and preschool theory of mind tasks reflect the EF demands of those tasks alone (the “expression alone” hypothesis; Kovács, 2009; Southgate, Senju, & Csibra, 2007). It is important to note, however, that not all expression accounts are mutually exclusive with construction accounts of theory of mind development or of the EF-Theory of Mind relationship. They merely argue that whenever the relevant understanding of beliefs does arise, it may fail to be expressed if EF skills are insufficient to meet the specific task demands of the probe for understanding.

Here we seek evidence in support of the basic hypothesis that the preschool measures of false belief understanding necessarily draw on executive function. We return in the general discussion to the stronger hypothesis that the developmental changes on theory of mind tasks observed in the preschool years may reflect improvements in EF *alone*.

Although the *prima facie* argument for a necessary role of EF in the expression of an understanding of false beliefs is compelling, there is no unequivocal evidence that EF is required to perform well on tests of false belief understanding. The observed developmental correlation between EF and belief understanding is obviously compatible with this hypothesis, but it is also compatible with other explanations, (e.g. Benson, Sabbagh, Carlson, & Zelazo, 2013; Frye et al., 1995; Lackner et al., 2010; Moses, 2001; Perner & Lang, 1999). Studies that attempt to reduce the EF skills required to pass tests of belief understanding by making reality less salient or by eliciting fewer prepotent responses often find better performance amongst 3-year-old children (e.g. Carlson et al., 1998; Wellman & Bartsch, 1988), but task changes intended to lessen EF demands may introduce other differences as well. Training studies aimed at improving children's EF skills also benefit belief understanding (Kloo & Perner, 2003), but such studies often take place over the course of several weeks or months, allowing for the possibility that improved EF skills contribute to the development of belief understanding in ways that go beyond greater capacity for expression.

Other findings challenge the view that an existing understanding of beliefs is simply unmasked as soon as preschoolers develop the requisite level of EF ability. For example, cross-cultural research has found that while Chinese preschoolers outperform American preschoolers on measures of EF they do no better on tests of belief understanding, and microgenetic research has shown that improvements in EF do not immediately extend to improvements in belief understanding (Sabbagh, Xu, Carlson, Moses, &

Lee, 2006; Flynn, 2007). These and other findings have led some authors to argue that the correlation between EF and belief understanding reflects a need for EF not in an expression capacity but rather in the very construction of an understanding of beliefs (Flynn, 2007; Moses, Carlson, & Sabbagh, 2004; Moses & Tahiroglu, 2010; Sabbagh et al., 2006; Wellman et al., 2001). This shift may be premature. While such results do militate against the hypothesis that improvement on preschool ToM tasks reflects improvements in EF *alone*, they are not inconsistent with an expression account of the relationship between EF and belief understanding wherein a certain level of EF is necessary, even if not sufficient, for success on the preschool tasks. Here, we propose to test the expression account directly by using the EF depletion paradigm.

1.3. The nature of EF depletion

There are now numerous demonstrations of EF, or “ego”, depletion in adults (see Hagger et al., 2010 for a review, though see also Hagger et al. (2016), for a recent failure to observe performance decrements following a specific depletion task in a large-scale pre-registered replication project), but the current studies represent the first direct tests for this phenomenon in children’s behavior. The results of these studies are relevant to debates about both the robustness of EF depletion and the exact mechanism behind the phenomenon seen in adults. With respect to robustness, a review of the literature shows that publication bias has likely resulted in the underreporting of studies of EF depletion that find null results (Carter, Kofler, Forster, & McCullough, 2015). Meta-analyses using a variety of statistical strategies to correct for this bias have come to different conclusions regarding the true effect size of EF depletion manipulations, ranging from no effect at all to a moderate effect of $d = 0.55$ (Carter et al., 2015; Inzlicht, Gervais, & Berkman, 2015). Here we report all experimental data we collected on EF depletion in children 4 years of age and older, and thus provide evidence for the presence or absence of depletion in preschool-aged children that is unbiased by a failure to include null results.

With respect to mechanism, there is a debate as to whether EF depletion effects reflect the reduction of a physical resource or whether they reflect the influence of mental content, such as theories about the mechanics of mental effort, or cognitive processes, such as calculations of opportunity cost. Baumeister and colleagues hypothesize that intense exercise of “will power” (or executive function) depletes a renewable physical resource, potentially glucose, which then needs to be replenished (Baumeister, Vohs, & Tice, 2007; Gailliot et al., 2007). If this is the correct interpretation of the phenomenon, EF depletion effects should be observable at any age. However, the evidence that self-control lowers blood glucose levels has been called into question by reanalyses and failures to replicate (Kurzban, 2010; Molden et al., 2012). Moreover, several researchers have pointed out that there is no evidence that other processes that rely on intense neural activity, such as visual perception, result in depletion effects on either performance or blood glucose (Beedie & Lane, 2012; Kurzban, Duckworth, Kable, & Myers, 2013).

Other accounts of EF depletion appeal to reductions in the motivation to engage in cognitive control stemming from sources such as opportunity cost calculations comparing the value of the task at hand to other options (Kurzban et al., 2013; Muraven & Slessareva, 2003) or a drive to balance cognitive control with “cognitive leisure” or the pursuit of immediate gratification (Inzlicht & Schmeichel, 2012; Kool & Botvinick, 2014). Yet others invoke the influence of intuitive theories of mental effort on performance (Job, Dweck, & Walton, 2010). All of these accounts require some degree of capacity for metacognitive monitoring and control,

capacities that are slow to develop and still maturing during the elementary school years (Flavell, 1979). Preschool-aged children do show some ability to monitor cognitive states such as feelings of uncertainty and judgments of learning (Lyons & Ghetti, 2010), but they often fail to use this monitoring to strategically control their behavior (Destan, Hembacher, Ghetti, & Roebbers, 2014; Schneider & Lockl, 2008). Young children are also unlikely to hold theories about hard intellectual work or consciously reflect on the value of task performance (Flavell, Green, Flavell, Harris, & Astington, 1995). Thus the observation of EF depletion in preschool-aged children would place limits on the metacognitive capacities that can be plausibly hypothesized to underlie the depletion phenomenon.

1.4. The current experiments

Experiments 1 and 3 test for evidence of a depletion effect in children, and Experiments 2 and 4 use depletion manipulations as a methodological tool to investigate the role of EF in belief understanding in 4- and 5-year-old children. Children this age generally do well on false belief tasks, so testing how EF depletion impacts performance on both a standard prediction version of this task as well as an explanation version will help us to tell whether EF is necessary for the expression of belief understanding and, if so, what aspects of false belief reasoning give rise to this need.

2. Experiment 1

Experiment 1 assessed the effects of EF depletion on children’s performance on a conflict inhibitory control task in which they had to ignore salient response options in order to make non-canonical responses, such as saying “circle” when looking at a square.

2.1. Materials and methods

2.1.1. Participants

Thirty-two 5-year-old children participated (17 males; ages 60 months 12 days to 71 months 11 days). This sample size was decided upon in advance, and was similar to but somewhat smaller than sample sizes in adult studies of EF depletion, as the use of a within- rather than between-subjects design afforded greater power to detect a potential change in EF performance following the depletion manipulation. Three additional children were excluded due to a failure to follow task instructions.

2.1.2. Procedure

The study began with one of two tasks (shape naming or direction naming) that provided a baseline measure of inhibitory control for each participant. For the first block of 18 trials participants were instructed to respond with congruent labels for pictures presented on a deck of cards (e.g. to say “up” when they saw an upward pointing arrow and “down” when they saw a downward pointing arrow). Immediately afterward they completed a second block of 18 trials where they were asked to respond with the incongruent label for each card (e.g. to say “down” when they saw an upward pointing arrow and “up” when they saw a downward pointing arrow), requiring the inhibition of a default, and practiced, response. Before each block, participants were encouraged to respond as quickly as they could while still giving the correct answers. When a participant made an error during the task the experimenter paused and, if the child did not self-correct, pointed out the correct response.

Next, children participated in either an EF depletion event or a filler event. In the depletion event, each participant was shown an opaque box and was told that it contained a large number of toys

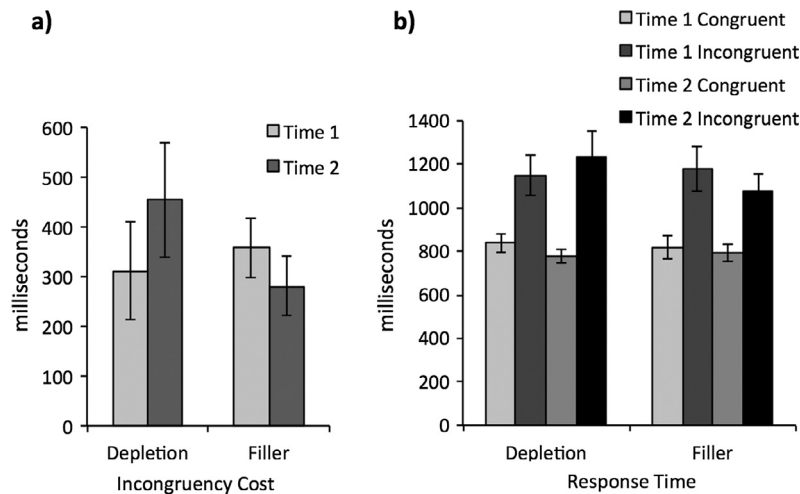


Fig. 1. Data from Experiment 1, including (a) incongruity costs (mean incongruent RT – mean congruent RT) for inhibitory control tasks conducted at Time 1 and Time 2 (before and after the depletion or filler event), and (b) mean RT for both the congruent and incongruent blocks of trials at each time point. Error bars represent standard error of the mean.

from which they could choose one to keep, but only if they waited in the testing room for several minutes alone with the box and did not open it, requiring the child to engage in extended delay of gratification. The experimenter confirmed that the child was willing to participate in this event and then left the room for 5 min, watching the participant via a hidden camera to make sure he or she did not look inside the box. The experimenter then returned and encouraged the participant to open the box of toys and choose a toy to keep. The filler event consisted of an introduction to the same box, its immediate opening, and a 5-min period in which the participant and experimenter played with the toys. These participants were also encouraged to choose a toy to keep. Thus, the delay between the two measurements of EF, the presence of the box of toys, and the choice of a toy to keep were constant across the two conditions.

Finally, participants completed another round of the inhibitory control task described above. Participants who previously completed the direction version of the task now completed the shape version, and vice versa.

2.1.3. Data analysis

The reaction time (RT) for each trial of the inhibitory control tasks was defined as the duration between the placement of each card on the table and the beginning of the participant's response. RTs were coded offline by a researcher blind to the condition assigned to each participant. A second researcher coded 25% of participants, and correlation between the two coders was $r = 0.99$. The first trial of each block was excluded as a practice trial, and reaction times for the remaining trials on which participants responded correctly were averaged to create a mean response time for the congruent and incongruent blocks of each task for each participant. We calculated an incongruity cost for each task by subtracting the mean response time for congruent trials from that for incongruent trials, a procedure that helps correct for heterogeneity in basic response times across participants (Coulthard, Nachev, & Husain, 2008). A lower incongruity cost reflects more effective application of EF skills.

2.2. Results and discussion

Error rates averaged less than 10% across participants for each block in each administration of the task. A repeated measures ANOVA examined the effect of within-subjects factors of task time (before vs. after depletion/filler manipulation) and trial type (con-

gruent vs. incongruent) and the between-subjects factors of condition (depletion vs. filler), task order (shape or direction first), and gender on error rates. There was a significant effect of trial type ($F(1, 24) = 14.98, p < 0.01$), with participants making more errors on incongruent trials ($M = 7.5\%$) than congruent trials ($M = 3.7\%$). There were no significant main effects or interactions involving task time or condition.

Next we investigated the effect of our experimental variables on reaction times, analyzing only those trials on which participants made a correct response. Participants were faster to respond on congruent than incongruent trials for both the shape version (812 ms vs. 1216 ms, $t(31) = 5.68, p < 0.001$) and the direction version (810 ms vs. 1101 ms, $t(31) = 6.15, p < 0.001$) of the task. We analyzed the effects of our manipulations on RT in two steps. First, after the incongruity costs were calculated for each task (Fig. 1a), a repeated-measures ANOVA examined the effects of the within-subjects factor of task time (before or after the intervening depletion/filler event) and the between-subjects factors of condition (depletion or filler), task order (shape or direction version first), and gender on incongruity costs. If children, like adults, utilize EF capacities less effectively immediately after having applied them to another task, then children in the depletion condition should show a greater increase in incongruity cost from pretest to posttest than do children in the filler condition. Using the interaction between condition and task time to test for the presence or absence of a depletion effect controls for any positive or negative influence of general factors such as practice or experiment duration on incongruity cost that would have affected the filler group as well as the depletion group. There was indeed a significant interaction between task time and condition ($F(1, 24) = 4.38, p < 0.05$). There were no other main effects or interactions involving these variables. As can be seen from Fig. 1a, participants in the filler condition displayed a slightly lower incongruity cost on the second EF task ($M = 281$ ms) than on the first ($M = 358$ ms), though the difference was not significant ($t(15) = 0.99, p = 0.34$). In contrast, participants in the depletion condition displayed a higher incongruity cost on the second task ($M = 454$ ms) than on the first ($M = 312$ ms), with a trend toward significance ($t(15) = 1.65, p = 0.06, 1$ -tailed¹). The important result is the interaction, which shows that the EF depletion manipulation led to a significantly greater increase in incongruity cost for the

¹ A one-tailed prediction is justified by the adult literature on EF depletion (Hagger et al., 2010).

depletion participants compared to the control participants, reflecting a relative reduction in available EF capacity.

A second analysis explored the source of this interaction further, asking whether condition and task time affect both congruent and incongruent trials. The depletion hypothesis predicts a selective effect on incongruent trials, but absolute incongruity cost could have risen due to a proportional slowing of all trials. As can be seen from Fig. 1b, the depletion task did not cause participants to slow down on all trials; depletion participants' average RT on congruent trials was slightly, though not significantly, faster at Time 2 ($M = 778$ ms) compared to Time 1 ($M = 838$ ms; $t(15) = 1.31$, $p > 0.2$). These values are virtually indistinguishable from those for the filler condition (Time 2 $M = 793$ ms; Time 1 $M = 818$ ms). A repeated measures ANOVA examining the effects of condition and test time on RTs to congruent trials revealed no main effects or interactions (all $p > 0.1$); participants' performance on the congruent trials was remarkably stable. Thus, the larger increase in depletion participants' incongruity cost from Time 1 to Time 2, relative to that of filler participants', was not a product of proportional slowing down on all trials, but rather a selective slowing on the incongruent trials.

Experiment 1 is the first demonstration of EF depletion in children, and adds to the evidence for the existence of the phenomenon. Furthermore, that 5-year-old children exhibit it suggests that intuitive theories concerning mental will power and sophisticated metacognitive control are not the sole loci of the phenomenon. Having established that a depletion event effectively decreased participants' ability or proclivity to exercise EF on a subsequent task, Experiment 2 explores whether depleting EF interferes with 5-year-olds' performance on a false belief task.

3. Experiment 2

Experiment 2 contrasted the performance of children in depletion and filler conditions on a subsequent false belief task in order to see whether a reduction in available EF resources would interfere with the expression of an otherwise functional understanding of beliefs. To test whether any observed effect of EF depletion is specific to reasoning about false beliefs, the task included a control scenario that was similar in complexity but did not involve a false belief.

3.1. Materials and methods

3.1.1. Participants

Seventy 5-year-old children participated (33 male; ages 60 months 7 days to 71 months 28 days), 35 in the EF depletion condition and 35 in the filler condition. This sample size, chosen in advance, was larger than in Experiment 1 and more similar to past studies on adult EF depletion in order to compensate for a change to a between-subjects design for key measures of performance. Children were only tested on their prediction of others' behavior in false belief and control scenarios after the depletion manipulation, and thus we had no baseline measure of belief understanding to help remove variability due to individual differences apart from those induced by EF depletion. Six additional participants were excluded, 4 because they left the testing room before the end of the depletion event and 2 because their answers for the control stories were outliers (see Data Analysis below).

3.1.2. Procedure

The experiment began with either an EF depletion event or filler event, conducted in the same manner as the manipulation events in Experiment 1. Next, all participants completed a change-of-location false belief task designed to allow for a continuous mea-

sure of accuracy (Bernstein, Thornton, & Sommerville, 2011; Sommerville, Bernstein, & Meltzoff, 2013). For 31 participants the same experimenter administered both the EF depletion or filler event and the false belief task, while for the remaining 39 participants the false belief task was administered by a second experimenter blind to whether the child was in the depletion or filler condition. For the false belief task, the participant and experimenter were seated on opposite sides of a $5' \times 1.5' \times 1'$ box full of Styrofoam packing peanuts. The experimenter told the participant three stories. Each story featured a new protagonist who hid an object somewhere in the Styrofoam, which represented a different substance in each story (i.e., snow, ice, and bubbles), and then left. In the two false belief stories, a second character moved the object to a new location in the Styrofoam without telling the protagonist. In the "2-object" control story, the second character had a different object, and hid that one in the Styrofoam rather than moving the original object. This story thus matched the false belief story in terms of characters and locations attended to. In all stories the two relevant locations were between 14 and 24 in. apart and were marked before the study with tacks on the back of the box that were visible to the experimenter but not the participant.

To prevent visual fixation on any of the locations involved in the stories, after the experimenter finished each story, she got out an "I Spy" book and engaged the participant in a visual search game for approximately 1 min (Sommerville et al., 2013). The experimenter then told the participant that the protagonist was coming back to look for the object he hid, and asked the children to point to the location in the box where the protagonist would look. A ruler was lined up from the participant's finger to the back of the box, and the experimenter used a tack to mark the location of each response.

3.1.3. Data analysis

On several occasions children seemed to forget both the first and second hiding places, pointing to a spot in the box far from either location. We calculated error scores comprising the distance between each answer and the closest hiding location, regardless of whether it constituted the correct or incorrect answer for the task, and then identified outliers more than two standard deviations from the mean error. This analysis resulted in the exclusion of three false belief answers (two from the depletion condition, one from the filler condition) and two 2-object answers (both from the depletion condition), accounting for 2.4% of all trials. The children who produced outliers on the 2-object story could not be included in the data set (as there was only one such story), and so were replaced. Children who produced an outlier answer on a false belief story were included using only their score from the other false belief story.

Given the novelty of this continuous version of the false belief task, data were coded and analyzed two ways. First, for categorical coding, each response was coded as correct or incorrect based on whether the child pointed closer to the correct location or to the distractor location. The proportions correct were calculated for false belief stories (values per child of 0, 0.5, or 1) and for 2-object stories (0 or 1) and are displayed in Fig. 2a.

Second, for proportional bias coding, the distance between the participant's answer and the correct location was divided by the total distance between the two hiding locations. Answers that deviated in the direction of the distractor location were given positive values while answers that erred in the opposite direction were given negative values to indicate biases toward or away from the distractor location. Bias scores for the two false belief stories were averaged and are displayed in Fig. 2b, along with the bias scores for the two-object control story.

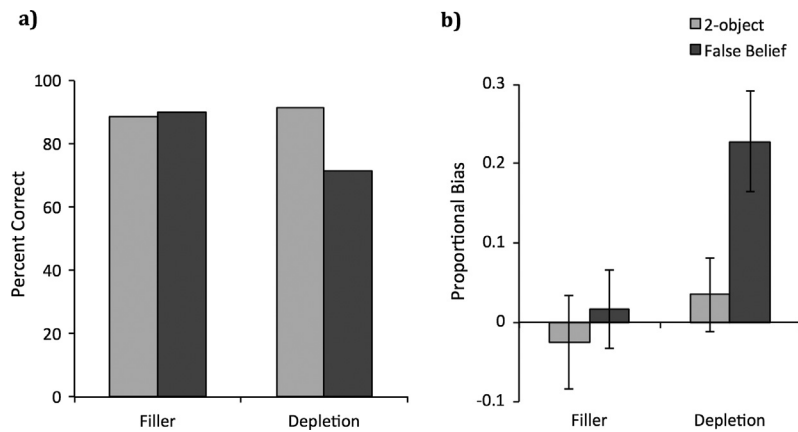


Fig. 2. Results from Experiment 2 for both 2-object control stories and false belief stories coded (a) categorically as percent correct and (b) continuously as a measure of bias in response location. In panel b, a bias score of 0 represents a correct answer, with positive and negative values representing deviations toward and away from the distractor location, respectively. Error bars represent standard error of the means.

A preliminary analysis examined the effect of whether the false belief experimenter was or was not blind to condition. There were no main effects or interactions involving this variable, so subsequent analyses collapsed across it.

3.2. Results and discussion

As seen in Fig. 2a, the categorical coding of the data demonstrated that children in the filler condition performed equally well on the 2-object control stories (89%) and the false belief (90%) stories, as expected for 5-year-olds who robustly succeed at standard false belief tasks (Wellman et al., 2001). In contrast, participants in the depletion condition performed better on the 2-object (91%) than on the false belief (71%) stories. To establish whether the difference between the two conditions was statistically reliable, the proportion correct for false belief stories was subtracted from that for 2-object stories for each participant, resulting in difference scores that ranged from 1 (passing 2-object but not false belief stories) to -1 (passing false belief stories but not 2-object story). The difference scores for depletion and filler participants were compared using a Mann-Whitney *U* test. The Mann-Whitney revealed that the difference scores varied reliably between the filler condition ($M = -0.01$) and the depletion condition ($M = 0.20$; $U(69) = 811$, $Z = -2.33$, $p < 0.05$). As Fig. 2a reveals, this result is due entirely to relatively poor performance on the false belief stories by children in the depletion condition, a conclusion that is confirmed in the analyses of proportional bias (see Fig. 2b, and below).

A repeated measures ANOVA examined the effects of condition (depletion or filler), story type (false belief or 2-object), order of story type, and gender on the bias scores. There was a significant main effect of story type ($F(1,58) = 5.98$, $p < 0.05$). Participants had a mean bias towards the incorrect location for the false belief stories ($M = 0.12$) and none for the 2-object control stories ($M = 0.01$). There was also a trend toward an interaction between story type and condition ($F(1,58) = 2.69$, $p = 0.1$). Examining the means revealed that depletion participants showed significantly greater proportional bias on false belief ($M = 0.23$) than control stories ($M = 0.04$; $t(34) = 2.68$, $p < 0.05$). They also showed greater proportional bias on false belief stories than filler participants did ($M = 0.02$, $t(68) = 2.64$, $p < 0.05$). Participants in the filler condition showed virtually no bias on either type of story (see Fig. 2b). No other main effects or interactions approached significance. Thus, both the categorical and proportional bias coding support the conclusion that EF depletion selectively impairs performance on false belief stories, with the strongest evidence coming from the categor-

ical coding.² These results demonstrate that even at 5 years of age, when children are clearly in possession of an explicit understanding of beliefs, they still rely on EF resources to put that understanding to use in predicting protagonists' actions based on those false beliefs.

4. Experiments 3a and 3b

In the adult literature many different depleting tasks, drawing on different aspects of EF, have been shown to affect performance on subsequent measures of EF. In Experiment 3a we explored the generality of EF depletion effects in preschool children with a new group of 5-year-olds, using a "Go/No Go" task in place of the delay of gratification task as our depleting measure. In Experiment 3b, we further investigated the developmental course of EF depletion by testing 4-year-olds with the same method.

4.1. Experiment 3a. Materials and methods

4.1.1. Participants

Fifty-six 5-year-olds (25 male; ages 60 months 3 days to 71 months 23 days) participated, split into filler and depletion groups. This sample size was based on a power analysis of the data from Experiment 1 and selected to provide an 80% chance of replicating the differing changes in EF performance for the depletion vs. filler groups. An additional four children participated but were excluded from the final dataset because they failed to follow the rules of the naming game.

4.1.2. Procedure

The overall design was the same as in Experiment 1, with two sets of congruent and incongruent naming tasks separated by either a depletion or filler task. The up and down direction naming task was replaced by a face naming task in which participants first had to correctly label happy and sad faces, followed by an incongruent naming block in which they had to respond with "sad" for a happy face and "happy" for sad faces. The shape naming task was kept the same, and order of shape and face naming tasks was again counterbalanced across participants.

² The pattern of bias scores observed here failed to conform to the previous report of 5-year-olds' performance on a similar continuous false belief task. Unlike Sommerville et al. (2013), we found no average bias toward the incorrect location on false belief trials for non-depleted participants. The distribution of responses for depleted children, for whom significant average bias was observed, was strongly bimodal, also contradicting the original report. Thus, coding distance from the correct answer, rather than categorically binning responses as correct or incorrect, likely did more to increase noise related to imprecise memories for object locations than to increase sensitivity to failures of belief attribution.

After one experimenter had administered the first set of naming tasks, she left and a second experimenter came into administer the depletion or filler task. Participants in the depletion condition completed a Go/No-Go task in which they were seated in front of a laptop and shown a series of trials featuring either a blue square or a red circle. Participants were told that when presented with the square they should hit the blue square symbol affixed to the table near their right hand (a “Go” trial) and that when presented with the red circle they should refrain from hitting either the blue square or the red circle symbol affixed near their left hand (a “No-Go” trial). After participants could successfully complete several practice trials according to these rules, they completed two rounds of this task, each consisting of 20 trials that lasted 3 s apiece. The first block consisted of 65% Go trials and 35% No-Go trials, and the second block consisted of 75% Go trials and 25% No-Go trials. The relative prevalence of Go trials contributes to the difficulty of inhibiting responses on No-Go trials, generating the demand for inhibitory control designed to draw on children’s EF. Meanwhile, the experimenter recorded the participant’s performance on a second computer via a scale that was visible to the participant. A response was counted as incorrect if the participant hit either of the two symbols when the red circle appeared onscreen or if the participant failed to hit the blue square while a blue square was onscreen. If the participant completed all 20 trials in a block correctly the scale filled to the top and was then replaced by a graphic of fireworks to celebrate the participant’s performance, but no verbal feedback was given during the block.

The filler task was similar in setup and duration, but instead of seeing blue squares and red circles, participants saw only blue squares positioned either on the right or the left of the screen and had two blue square response options on the left and the right. Filler participants were asked to respond by hitting the response option on the same side as the onscreen square on each trial, akin to congruent trials in a Simon task (e.g. Davidson, Amso, Anderson, & Diamond, 2006; Shaffer, 1965).

After the second experimenter had completed either the depletion or filler task with the participant, she left the room and the first experimenter returned to administer the second congruent and incongruent naming task.

4.1.3. Data analysis

RTs for each trial of the two naming tasks were coded in the same manner as in Experiment 1. Twenty-three percent of participants were coded by two experimenters, and the correlation between their measures of RTs was $r = 0.92$. We again calculated the average RT for correct trials in each congruent and incongruent block, excluding the first trial. For this data set, we also eliminated outlier trials for which RTs were more than 3 standard deviations from the mean RT for the block, which mainly eliminated trials in which the participants disengaged from the task and needed to be reminded to respond by the experimenter, and a small number of trials with RTs less than 300 ms likely due to anticipatory responding (Davidson et al., 2006).³ We calculated each partici-

³ In Experiment 1 no RTs under 300 ms were observed. Furthermore, if the experimenter needed to prompt a response, she asked, “What is this called in the [name/opposite] game?” Since this could have helped participants provide the correct answer, these trials were classified as experimenter errors and excluded from all analyses in Experiment 1, resulting in few aberrantly long RTs to correct responses (e.g. only 3 trials with RT > 5000 ms, all from the same participant). In Experiments 3a and 3b, experimenters used neutral language to prompt participants if necessary, resulting in an increase of technically correct responses recorded at excessively long RTs. The removal of outliers served to eliminate trials characterized by disruptions (e.g. a child’s attempt to begin a conversation with the experimenter mid-trial) that likely did not index the operation of the EF processes of interest. The number of trials excluded due to these two criteria was small, accounting for <1% of all trials across Experiments 3a and 3b.

part’s incongruency cost for each of the two naming tasks by subtracting the mean RT for congruent trials from the mean RT for incongruent trials (Fig. 3a). We then compared incongruency costs before and after the depletion or filler task to test whether these manipulations had differing effects on EF performance. As in Experiment 1, we then followed up with analyses of the factors that affected RTs on congruent trials separately.

4.2. Results and discussion

Error rates were low, as in Experiment 1. An ANOVA examining the effects of task time (pretest vs. posttest), trial type (congruent vs. incongruent) and condition (depletion vs. filler) found a main effect of trial type ($F(1,48) = 23.39$; $p < 0.001$), reflecting a higher error rate on incongruent trials ($M = 7.3\%$) than on congruent trials ($M = 3.7\%$). There was also a main effect of time ($F(1,48) = 14.28$; $p < 0.001$), reflecting a higher error rate for the blocks conducted before the filler or depletion manipulation ($M = 6.9\%$) than those conducted after the manipulation ($M = 4.2\%$), suggesting a practice effect. There were no main effects or interactions involving the condition (depletion vs. filler). Participating in the go-no go task did not differentially affect children’s tendency to err on either congruent or incongruent trials, relative to participating in the filler task.

As in Experiment 1, we then analyzed reaction times for correct naming trials. Participants were again faster to respond on congruent than incongruent trials of both naming tasks (shape naming: congruent $M = 655$ ms, incongruent $M = 1156$ ms, $t(55) = 14.16$, $p < 0.0001$; face naming: congruent $M = 752$ ms, incongruent $M = 1314$ ms, $t(55) = 19.21$, $p < 0.0001$). Next we asked whether the participants in the depletion condition showed a greater increase in incongruency cost between pretest and posttest than did those in the filler condition. A repeated-measures ANOVA examined the effects of within-subjects factor of task time (pretest vs. posttest) and the between-subjects factor of condition (depletion vs. filler) on incongruency cost (see Fig. 3a). As in Experiment 1, there was a significant interaction between time and condition ($F(1,52) = 4.53$, $p < 0.05$). This reflects the fact that depletion participants displayed significantly higher incongruency costs after depletion ($M = 628$ ms) than before ($M = 493$ ms; $t(27) = 2.37$, $p < 0.05$), while incongruency costs for filler participants did not change appreciably from before ($M = 517$) to after ($M = 487$) the manipulation ($t(27) = 0.61$, $p > 0.5$). The difference in incongruency cost between the two groups of participants at Time 2, following the depletion or filler manipulations, was also significant ($t(54) = 2.02$, $p < 0.05$).

Fig. 3b displays the RTs for each type of trial (congruent vs. incongruent) at pretest and posttest for participants in the depletion and the filler conditions. Again, there were no differences between pretest and posttest RTs for congruent trials in either condition, though participants in the filler group were faster than the depletion group on congruent trials overall ($F(1,54) = 5.63$, $p < 0.05$). Rather, the increase in incongruency costs in the depletion condition were driven by longer reaction times for incongruent trials at posttest than at pretest.

Experiment 3a finds that engaging 5-year-olds in just a few minutes of a demanding response inhibition task interferes with deployment of their EF capacities in a response conflict task immediately thereafter. This replicates the pattern of results from Experiment 1, extending them to a different depletion manipulation. Just as with adults, different tasks, taxing different components of EF, yield depletion effects (Schmeichel, 2007).

The two experiments together show EF depletion to be a robust phenomenon in 5-year-old children. Experiments 1 and 3a are the only two attempts we have made to find depletion effects at this age. There is no file-drawer problem here; this work involved no

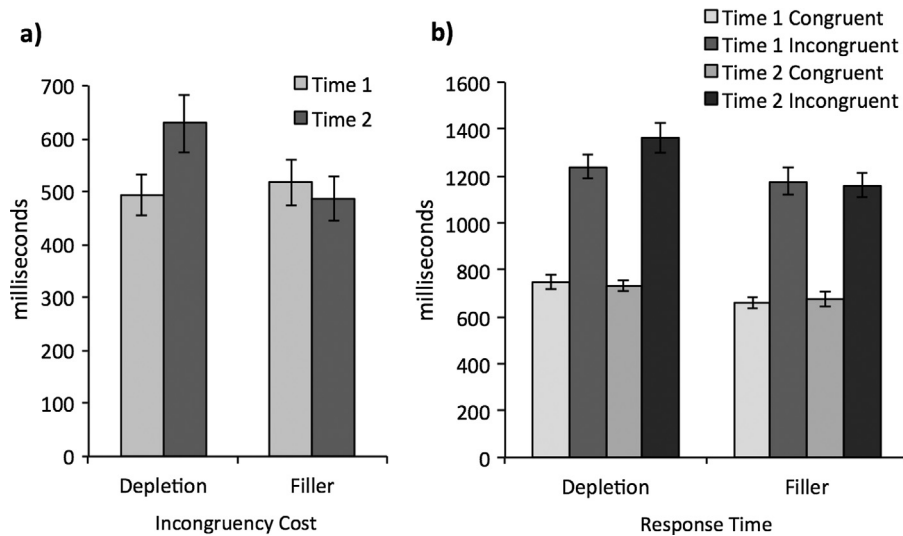


Fig. 3. Data from Experiment 3a, including (a) incongruity costs for inhibitory control tasks conducted at Time 1 and Time 2 (before and after the depletion or filler task), and (b) mean RT for both the congruent and incongruent blocks of trials at each time point. Error bars represent standard error of the mean.

extended piloting phase that yielded null results. As children this young have weak metacognitive control (Schneider & Lockl, 2008) and are likely to have few intuitive beliefs about mental effort (Flavell et al., 1995), these findings suggest that the mechanisms underlying EF depletion depend, at least in part, on processes that do not require sophisticated metacognition. If so, even younger children may be subject to depletion effects. Experiment 3b explores this issue, repeating Experiment 3a with 4-year-olds.

4.3. Experiment 3b. Materials and methods

4.3.1 Participants

Fifty-six 4-year-olds (23 male; ages 48 months 1 day to 59 months 25 days) participated, split into filler and depletion groups. An additional five children participated but were excluded from the final dataset because they failed to follow the rules of the naming game.

4.3.2 Procedure and data analysis

The procedure and data analysis were the same as in Experiment 3a. Twenty-five percent of participants were coded by two experimenters, and the correlation between their times was $r = 0.95$.

4.4. Experiment 3b results and discussion

An ANOVA comparing error rates for each block of trials for both participant groups before and after the manipulations showed a main effect of trial type ($F(1,48) = 11.28$; $p < 0.01$), reflecting a higher error rate on incongruent trials ($M = 6.9\%$) than on congruent trials ($M = 4.1\%$). There was also a main effect of task time ($F(1,48) = 4.99$; $p < 0.05$), with participants making more errors during the first administration of the naming task ($M = 6.3\%$) than the second administration ($M = 4.7\%$). There were no interactions involving these variables, nor were there any effects of condition (depletion vs. filler).

Participants were faster to respond on congruent than incongruent trials of both naming tasks (shape naming: congruent $M = 766$ ms, incongruent $M = 1311$ ms, $t(55) = 9.09$, $p < 0.0001$; face naming: congruent $M = 850$ ms, incongruent $M = 1449$ ms, $t(55) = 13.01$, $p < 0.0001$). A repeated-measures ANOVA examining

the effect of the within-subjects factor of task time (pretest vs. posttest) and the between-subjects factors of condition (depletion vs. filler) on incongruity cost found no main effect of or interactions between these factors (Fig. 4). Incongruity cost was similar at Time 1 and Time 2 for both the depletion group (Time 1 $M = 522$ ms; Time 2 $M = 513$ ms) and the filler group (Time 1 $M = 663$ ms; Time 2 $M = 596$ ms). These data thus provide no evidence that 4-year-olds as a group are subject to EF depletion.

However, 4-year-olds' performance on the naming task was highly variable, with a standard deviation for incongruity cost at Time 1 (430 ms) that was twice as large as that for 5-year-olds (215 ms). This difference reflects a number of 4-year-olds with exceptionally high initial incongruity costs, resulting from congruent RTs that were similar to other participants' and incongruent RTs that were much higher. An exploratory examination of the data suggested that these children tended to show much lower incongruity costs at Time 2, reflecting either practice effects or regression to the mean. To assess whether the lack of depletion observed for the age group as a whole reflected the imperviousness of younger children to executive function depletion or the masking of such an effect by some children for whom repetition lead to

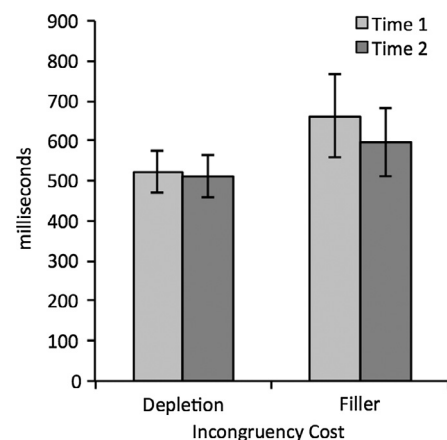


Fig. 4. Incongruity costs displayed by 4-year-olds in Experiment 3b during inhibitory control tasks performed before (Time 1) and after (Time 2) the depletion or filler manipulation. Error bars represent standard error of the mean.

large gains, we looked again at the relationship between incongruency cost and depletion in 4-year-olds whose initial performance was more similar to that of typical 5-year-olds. A maximum incongruency cost threshold of 800 ms at Time 1, which encompassed 95% of the 5-year-old sample, was applied to the 4-year-old sample. This criterion led to the exclusion of 14 4-year-olds (25% of the sample), evenly distributed across the conditions, who had indeed shown large decreases in mean incongruency cost from Time 1 ($M = 1149$ ms) to Time 2 ($M = 771$ ms).

Incongruency costs for the remaining 75% of 4-year-olds are displayed on Fig. 5a, as function of condition and test time. The pattern of incongruency cost changes matched that of 5-year-olds, with depletion participants displaying an increase in incongruency cost from Time 1 ($M = 394$ ms) to Time 2 ($M = 530$ ms), trending strongly toward significance ($t(21) = 1.95$, $p = 0.065$), despite the decrease in power resulting from the reduction in sample size. Filler participants' incongruency cost remained stable from Time 1 ($M = 415$ ms) to Time 2 ($M = 425$ ms; $t(19) = 0.14$, $p > 0.8$). Inspection of Fig. 5b shows this difference is due mainly to RTs on the incongruent trials, as changes in congruent RT were similar across condition. The performance of these selected 4-year-olds was statistically compared to that of the 5-year-olds in a repeated-measures ANOVA that assessed the effects of the within-subject factor of task time (pretest vs. posttest) and the between-subject factors of age group (5-year-olds vs. 4-year-olds) and condition (filler vs. depletion) on incongruency costs. The interaction between condition and task time was significant ($F(1,92) = 5.36$, $p < 0.05$), but there was no evidence that this interaction was affected by age ($p > 0.7$).

Four-year-olds have even less capacity for metacognitive monitoring and control than do 5-year-olds, and are less likely to possess metaconceptual beliefs about mental effort (Flavell et al., 1995; Lyons & Ghetti, 2013). That only a subset of 4-year-olds displayed an EF depletion effect raises the possibility that the phenomenon would be entirely absent in even younger children. This hypothesis merits further study, for persistent failures to observed depletion at ages 3 and 4 would certainly challenge the resource depletion models of the effects. However, we tentatively conclude that EF depletion *can* occur, under some circumstances, at age 4 as well as at age 5. Whether the effect of participating in an EF demanding task is observable in measures of children's

deployment of EF immediately thereafter depends upon characteristics of children's performance on the dependent measure of EF (in this case, incongruent trials on the shape/facial expression naming task). With a small to medium sample size, the effect of the depletion manipulation in a pretest/posttest comparison can easily be masked by large practice effects or individual differences on the dependent task. Further explorations of the developmental course of EF depletion effects will have to be sensitive to this fact.

5. Experiment 4

Having found that a variety of tasks can serve as effective EF depletion manipulations and may apply to a wider age range, we return to the question of the role EF plays in expressing belief understanding. Experiment 2 provided clear evidence that EF plays a role in expressing an understanding of false beliefs through predicting where a protagonist with a false belief will search. However, Experiment 2 does not shed any light on *why* predicting where the protagonist will look for the object necessitates EF resources. Many have assumed that response inhibition places the major burden on EF in this task. The structure of common false belief prediction tasks is likely to require inhibiting pointing to the salient, true location of the object the protagonist is looking for (Bloom & German, 2000; Carlson et al., 1998). It is possible that these sorts of task demands entirely explain the effect of EF depletion observed on false belief performance in Experiment 2.

However, some researchers have posited roles for EF that are more central to the actual attribution of a false belief. One such proposal appeals to the 'curse of knowledge', arguing that children (and adults) suffer from an egocentric bias that makes it difficult not to attribute their own knowledge to others (e.g. Birch & Bloom, 2003). A related conjecture, based on the finding that preschoolers also have a hard time reasoning about false signs and maps, is that it is difficult to assign false beliefs to other because beliefs are taken, by default, to represent the true state of the world (Leekam, Perner, Healey, & Sewell, 2008; Sabbagh, Moses, & Shiverick, 2006). If young children are committed to this aspect of the nature of beliefs, then EF resources might be needed to resolve a conflict to this commitment whenever they must attribute a false belief.

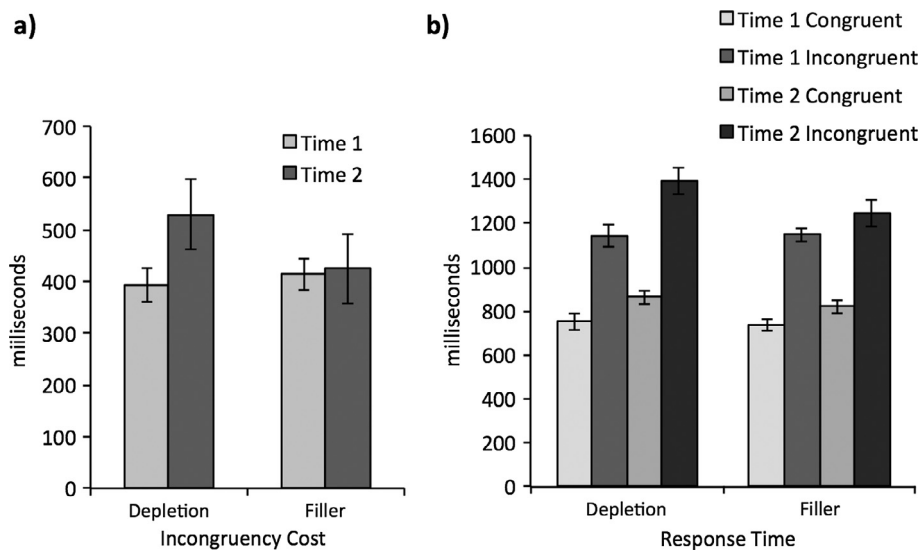


Fig. 5. Data from Experiment 3b for participants with Time 1 incongruency costs below 800 ms, showing (a) incongruency costs for inhibitory control tasks conducted at Time 1 and Time 2 (before and after the depletion or filler task), and (b) mean RT for both the congruent and incongruent blocks of trials at each time point. Error bars represent standard error of the mean.

If these latter hypotheses are correct and the attribution of false beliefs is difficult in and of itself, then EF resources may be necessary for successful belief reasoning even when EF demands on response production itself are reduced or eliminated. One way researchers have tried to reduce such demands is by asking children to explain behaviors guided by false beliefs rather than to predict them. In such explanation tasks, children are told a standard false belief story, but then they are told what the character does or says at the end of the story and are asked to explain this behavior (Bartsch & Wellman, 1989; Perner, Lang, & Kloo, 2002). This change means that the children no longer have to inhibit the urge to point to or respond with the salient wrong answer, but instead can focus on the cause of the misguided search. If the saliency of the true location was masking an accurate understanding of the character's beliefs, then children should now be able to express that understanding in their explanation. However, children do about equally well on prediction and explanation versions of false belief tasks, and performance on explanation tasks correlates with EF just as performance on prediction tasks does (Perner et al., 2002).

Assuming that the removal of task-specific inhibitory demands has eliminated the need for EF in *expression*, researchers have pointed to the relation of EF to successful explanation of a protagonist's search in an empty location as evidence that EF plays a role in children's *construction* of an understanding of beliefs (Moses & Tahiroglu, 2010; Perner et al., 2002). However, this assumption has never been directly tested. If EF plays a more central role in false belief attribution and is called upon regardless of the superficial nature of the task, it is possible that this correlation still reflects the role EF may play in the expression of the children's theory of mind even in these explanation tasks. If so, then depleting the child's EF resources should interfere with their performance on false belief explanation tasks as well as false belief prediction tasks. In Experiment 3 we sought to both replicate our finding that EF depletion impairs performance on prediction tasks and to test whether this effect extends to false belief explanation tasks as well.

5.1. Materials and methods

5.1.1. Participants

One hundred two 4- and 5-year-old children participated (ages 47 months 28 days to 71 months 12 days), 51 each in depletion and filler conditions. Data for this experiment were collected in the Discovery Center of the Boston Museum of Science. The sample size was the product of the aim to collect data from a similar number of participants as were recruited to participate in Experiment 2 over the course of semester-long research shifts at the museum; we ceased data collection at the end of the semester in which we surpassed 70 participants.

5.1.2. Procedure

The experiment began with either an EF depletion task or a matched filler task and then proceeded to the tests of false belief understanding. The depletion and filler manipulations consisted of the same Go/No-Go and matched filler tasks used in Experiment 3.

Subsequently, participants from both conditions completed the same two change-of-location false belief tasks with discrete object locations (as opposed to the continuous version from Experiment 2), administered as stories acted out with puppets and toys. The first story featured a single main character who put a toy in an initial location and was then absent while another character moved the toy to the second location. Like the false belief task in Experiment 2, this story ended with a prediction task where participants had to say where they thought the main character would look for the toy when he came back. The second story, modeled on those

used by Perner et al. (2002) in their study on false belief explanation, featured two characters who put a toy in one location together, but the toy was later moved by one of the characters while the other was absent. At the end of the story the two characters search for the toy simultaneously, and the participants were shown that the character who moved the toy looked for it where it was while the character who was absent during the movement looked in the now empty location. Participants were then asked why the latter character had looked in the location he did. If children gave an insufficient or wrong explanation or said they did not know the experimenter prompted them to elaborate or attempt another answer. The experimenter transcribed the participant's answer immediately after the task and, when parents agreed, responses were also audio recorded and used to verify or edit transcripts. No major edits affecting coding of responses were necessary.

5.1.3. Data analysis

Performance on the false belief prediction task was coded as correct if participants predicted that the main character would look for his toy where he left it and incorrect if they predicted he would look for it where it was at the end of the story. Performance on the false belief explanation task was based on participants' first unambiguous response and was coded as correct if participants referred to the character's mistaken belief (e.g. "he thinks it's in there") or if they referred to the events of the story that led to this belief (e.g. "that's where he left it"). Responses were coded as incorrect when participants referred to the status of the box (e.g. "because it isn't in there"), said they did not know why he looked there, or gave some other irrelevant response (e.g. telling the puppet where the toy was). Four children (one in the depletion condition, three in the filler condition) refused to make a verbal response to the explanation question, seemingly due to shyness, and were not given a score for the task. In addition to recording these scores separately, we also averaged them to give an overall false belief performance score of 0, 0.5 or 1. Given our strong a priori hypothesis regarding the effect of depletion on false belief reasoning (the results of Experiment 2 made it unlikely that EF depletion would result in *improved* false belief performance) and the low power of our binary measures of performance, we used one-tailed *z* tests for two proportions to compare scores for the depletion and filler groups on the overall average false belief performance. We then used a permutation test of participants' difference scores for the prediction and explanation scenarios to test whether there was any interaction between depletion condition and scenario type.

5.2. Results and discussion

The data are displayed on Fig. 6. We first asked whether the overall scores, averaged across both the prediction and explanation tasks, replicated our finding that EF depletion negatively impacts children's ability to reason about false beliefs. Children in the filler condition were correct 77.45% of the time, while children in the depletion condition were correct significantly less often, only 62.75% of the time ($z(100) = -1.621, p = 0.05, 1$ -tailed). To test for an interaction between condition and trial type we calculated the difference score for each participant between their prediction and explanation performance, ranging from -1 to 1 . We then performed a permutation test, shuffling the distributions of the difference scores 1000 times to see how one might expect such performance differences to be distributed across two random groups. The actual distribution of the difference scores across condition (depletion: 0.08, filler: 0.0625) was no different than would be expected by chance ($p > 0.9$), supporting the conclusion that the

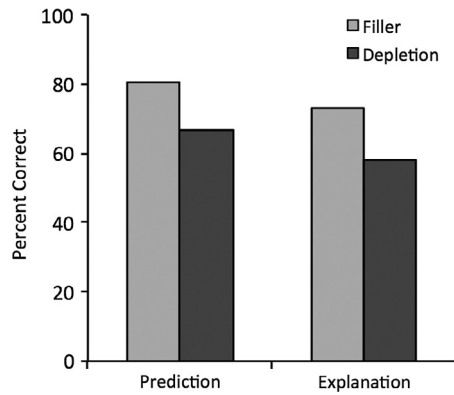


Fig. 6. Results of Experiment 4 showing percent correct on prediction- and explanation-based tests of false belief understanding in 4- and 5-year-olds that had previously completed either a depletion or filler task.

depletion condition had an equivalent effect on both the prediction and explanation tasks.⁴

These results replicate our finding from Experiment 2 that EF depletion significantly reduces children's ability to successfully reason about false beliefs. The fact that performance was slightly lower overall is likely a product of the wider age range used and the more distracting environment the study was conducted in. This lower performance may account for the fact that the absolute impact of depletion was slightly smaller than in Experiment 2; as a percentage of filler participants' false belief performance, the decrement resulting from depletion was relatively steady across experiments and task types (Experiment 2: 21%; Experiment 4, prediction: 17%; Experiment 4, explanation: 20%). Effect size may also have been impacted by the comparative brevity of the depletion task used here (3 min with intermittent demand for inhibitory control) compared to the one used in Experiment 2 (5 min of continuous inhibitory control).

In addition to replicating our previous findings, these results also show that the impact of depletion extends equally to both prediction tasks, which feature executive demands not specific to belief reasoning like ignoring a salient wrong answer, and explanation tasks, which are designed to limit these general EF demands while retaining the need to reason about beliefs. Although this particular experiment did not include a control condition devoid of EF demands to check for the specificity of depletion effects, the lack of impact of either depletion method on such control trials across the first three experiments lends support to the conclusion that EF is necessary for the expression of belief understanding across diverse types of assessment.

6. General discussion

Experiments 1 and 3 provide the first demonstration of EF depletion in preschool children and add to the evidence that the phenomenon of EF depletion does occur (Inzlicht et al., 2015). That depletion is found in preschoolers who lack metacognitive knowledge and strategies and are unlikely to hold explicit theories of mental effort (e.g., Flavell et al., 1995; Schneider & Lockl, 2008)

⁴ We also asked whether performance differed by condition within each of the two false belief trial types. On the prediction questions, participants in the filler condition were correct 80.39% of the time while children in the depletion condition were correct 66.67% of the time. This difference trended strongly toward but did not reach significance ($z(100) = -1.5204$, $p = 0.06$, 1-tailed). Performance on the explanation question showed a similar gap, with filler participants responding correctly on 72.92% of trials and depletion participants responding correctly on 58.00% of trials. Once again, this difference showed a strong trend toward significance ($z(96) = -1.5512$, $p = 0.06$, 1-tailed).

suggests that the phenomenon does not depend in full upon conscious, strategic decision making. These results are consistent with a resource model, or with a model in which subconscious calculations of opportunity cost or motivational shifts result in a relatively simple signal, such as the sensation of mental effort, that can be both monitored and acted upon by young children's relatively limited metacognitive capacities (Inzlicht & Schmeichel, 2012; Kurzban et al., 2013; Lyons & Ghetti, 2013).

As measured in Experiments 3a and 3b, the depletion effect was less robust among 4-year-olds than 5-year-olds, suggesting the need for follow-up studies with dependent measures better suited to 4-year-olds' EF capacities, as well as studies with still younger children. As Experiment 3b demonstrates, such exploration would need to carefully calibrate the difficulty of the pre- and post-tests that compare the performance of children who complete an EF depleting task with that of children in the control condition, so that the effect of practice on the task does not overwhelm the effect of the depletion manipulation. In addition, testing if observation of depletion effects correlates with children's ability to recognize a sensation of enhanced mental effort associated with depleting tasks could serve as a key test of the opportunity cost model of EF depletion (Kurzban et al., 2013).

That representations of false beliefs are selectively impaired by immediately prior exercise of EF (relative to a filler task—Experiments 2 and 4—and relative to representations of true beliefs—Experiment 2) simultaneously provides further evidence for EF depletion in preschoolers. In addition, Experiments 2 and 4 illustrate how depletion can be used to experimentally manipulate EF resources and, subsequently, test questions about the role of EF in various learning and reasoning processes. While many have speculated that EF skills play a necessary role in performance on measures of theory of mind used with preschool-aged children, the data presented here provide the first unequivocal support for the hypothesis that EF resources are needed to express an understanding of false beliefs. Five-year-olds generally possess a mature understanding of beliefs and the role that they play in guiding actions, as confirmed by participants in the filler conditions of Experiments 2 and 4. However, depleting EF resources, which can be done using a method like the one demonstrated in Experiment 1, masks this comprehension to a significant degree. These data demonstrate that possession of a folk psychological theory of beliefs is not, on its own, sufficient for success on tests of false belief understanding. In order to accurately use someone else's belief to make an explicit prediction about their behavior, substantial EF resources are also necessary.

The current research demonstrates a useful role for the EF depletion method in any domain. Task analyses aren't a foolproof way of assessing the role of EF in a cognitive process. For instance, researchers assumed false belief and false photograph tasks involved the same EF demands until it was shown that performance on the latter task does not actually correlate with EF performance (Sabbagh et al., 2006). In Experiment 4, we tested the assumption that explanation tasks do *not* place any substantial demands on EF resources. Many researchers have pointed out the surface features of standard false belief tasks that may add to EF demands, but hypotheses regarding the potential role of EF in central aspects of belief reasoning are less widely endorsed. The finding from Experiment 4 that EF depletion affects performance even on explanation tasks, which lack the usual surface demands such as a salient wrong answer, supports the view that EF is important for false belief attribution itself.

With these results, the hypothesis that development within theory of mind during the preschool years reflects changes in the on-line capacity to express already existing knowledge survives a critical test. While expression-based accounts can take several forms, the expression alone hypothesis holds that a lack of the EF

skills necessary to pass most explicit measures of theory of mind prevents preschoolers from expressing an understanding of beliefs present from infancy (Leslie, 1994). The data from Experiments 2 and 4 demonstrate that this sort of masking is possible.

That said, the present findings are also compatible with the possibility that substantial conceptual changes relevant to belief understanding occur in the fourth year of life and that these changes are necessary, in conjunction with the EF development that supports their expression, before preschoolers can begin to pass explicit false belief tasks. Performance on a variety of preschool ToM tasks is influenced by exposure to input illustrating the link between thoughts and behaviors, including input from explicit training (Appleton & Reddy, 1996; Slaughter & Gopnik, 1996), amount of mental state language in parental input (Ruffman, Slade, & Crowe, 2002), and environmental factors such as having older siblings (Perner, Ruffman, & Leekam, 1994). Finally, as mentioned above, Sabbagh et al. (2006) compared Chinese and American preschoolers on large batteries of EF measures and ToM measures. In this sample, the Chinese children were a full six months ahead of the American children on all of the EF measures, but the two populations were identical in their performance on the ToM tasks. Thus, superior EFs are not sufficient for superior performance on ToM tasks. This result undermines the claim that the development of EF can completely explain the developmental changes on performance on ToM tasks, contrary to the expression alone hypothesis. Some learning or construction specific to ToM is also implicated in the observed developmental changes in the preschool years.

However, that learning or construction plays a role in the developmental changes observed in the preschool years does not guarantee that EF is drawn upon in the learning process. It depends what that learning process is. Associative learning, for example, might not draw heavily on EF. It might be that EF plays a role in both the expression and construction of belief understanding, or only in the expression of belief understanding, even if that understanding undergoes substantial representational change in the preschool years. In one study providing empirical support for the conclusion that EF is drawn upon in the learning process as well, Benson et al. (2013) found that EF predicts which 3-year-olds improve on false belief tasks through training, concluding that EF is required for the construction of an explicit understanding of beliefs. This study does provide important evidence, but its conclusion must be tempered by several caveats. First, the improvements on the posttest were mainly restricted to the change of location false belief tasks on which children were trained; there were no effects on other tasks reflecting an explicit understanding of beliefs. Furthermore, the effect was evident after the very first training trial. Thus, this training intervention did not necessarily lead to the construction of an explicit, generalizable change in children's theory of mind. Indeed, Bartsch and Wellman's (1995) study of spontaneous speech locates 3 years, 0 months as the age at which children construct a representational theory of mind. Thus, an expression account of these results is also possible: children with stronger EF skills may have been better able to understand the feedback provided and to use that feedback to improve the expression of their belief understanding on particular tasks, including explanation tasks which Experiment 4 demonstrates are subject to the same EF-related performance constraints as prediction tasks.

As of now, there is no conclusive evidence either way as to what role EF plays in the processes that support the development of an explicit understanding of beliefs in the preschool years. Given that the current data show that the EF-belief understanding correlation could be explained by the need for EF skills in order to express one's concept of belief after it has been acquired, further evidence is required to support of the claim that EF plays an additional role

in the construction of a mature theory of mind. Additional EF depletion studies could provide relevant evidence. In training studies, for example, if EF depletion administered prior to training sessions but not the final test session reduced the benefit of training, we could potentially conclude that EF also plays a role in the construction of an understanding of beliefs. The same type of study design could be used in other domains where correlational studies have suggested a link between EF resources and learning but a causal connection has yet to be demonstrated (e.g. Baker et al., 2011; Bull & Scerif, 2001; Zaitchik et al., 2013).

Acknowledgements

The authors thank Meg Barrow, Julie Barzilay, Michael Chilazi, Cessna Mac, and Andrea Sem Stoltenberg for their assistance collecting data and Kate Hobbs, Mariko Moher, Amy Skerry and Elizabeth Spelke for commenting on drafts of the manuscript. This work was supported by the National Science Foundation (DRL-0940140).

References

- Appleton, M., & Reddy, V. (1996). Teaching three year-olds to pass false belief tests: a conversational approach. *Social Development, 5*, 275–291.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110–118.
- Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science, 14*, 270–279.
- Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946–964.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252–1256.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351–355.
- Beedie, C. J., & Lane, A. M. (2012). The role of glucose in self-control another look at the evidence and an alternative conceptualization. *Personality and Social Psychology Review, 16*, 143–153.
- Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*, 1615–1627.
- Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false-belief task. *Experimental Aging Research, 37*, 481–502.
- Birch, S. A. J., & Bloom, P. (2003). Children are cursed: an asymmetric bias in mental-state attribution. *Psychological Science, 14*, 283–286.
- 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.
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25–B31.
- 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.
- 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., & Hix, H. R. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.
- Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General, 144*, 796–815.
- Coulthard, E. J., Nachev, P., & Husain, M. (2008). Control over conflict during movement preparation: role of posterior parietal cortex. *Neuron, 58*, 144–157.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.
- Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568–570.
- Destan, N., Hembacher, E., Ghetti, S., & Roebbers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5-to 7-year-old children. *Journal of Experimental Child Psychology, 126*, 213–228.

- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, 318, 1387–1388.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *American Psychologist*, 34, 906–911.
- Flavell, J. H., Green, F. L., Flavell, E. R., Harris, P. L., & Astington, J. W. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*, 60, i–113.
- Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, 16, 53–59.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., ... Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92, 325–336.
- German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': evidence from compromised belief–desire reasoning in old age. *Cognition*, 101, 129–152.
- Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A., ... Zwienerberg, M. (2016). A multi-lab pre-registered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573.
- Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychological Bulletin*, 136, 495–525.
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16(2), 233–253.
- Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, 16, 174–180.
- Inzlicht, M., Gervais, W., & Berkman, E. (2015). Bias-correction techniques alone cannot determine whether ego depletion is different from zero commentary on Carter Koffler Forster & McCullough. Retrieved October from social science research network <http://ssrn.com/abstract=2659409>.
- Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, 7(5), 450–463.
- Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, 21, 1686–1693.
- Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development*, 74, 1823–1839.
- Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, 143, 131–141.
- Kovács, A. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12, 48–54.
- Kurzban, R. (2010). Does the brain consume additional glucose during self-control tasks? *Evolutionary Psychology*, 8, 244–259.
- Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, 36, 661–679.
- Lackner, C. L., Bowman, L. C., & Sabbagh, M. A. (2010). Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia*, 48, 1767–1774.
- Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology*, 26, 485–497.
- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 193–200.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45–85.
- Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: two conjectures. *Developmental Science*, 1, 247–253.
- Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development*, 84, 726–736.
- Lyons, K. E., & Ghetti, S. (2010). Metacognitive development in early childhood: new questions about old assumptions. In A. Efklides & P. Misaillidi (Eds.), *Trends and Prospects in Metacognition Research* (pp. 259–278). New York, NY: Springer.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Molden, D. C., Hui, C. M., Scholer, A. A., Meier, B. P., Noreen, E. E., D'Agostino, P. R., & Martin, V. (2012). Motivational versus metabolic effects of carbohydrates on self-control. *Psychological Science*, 23, 1137–1144.
- Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development*, 72(3), 688–690.
- Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2004). On the specificity of the relation between executive function and children's theories of mind. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 131–145). Mahwah, NJ: Erlbaum.
- Moses, L. J., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theories of mind. In B. Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds.), *Self- and social-regulation: exploring the relations between social interaction, social cognition, and the development of executive functions* (pp. 218–233). New York, NY: Oxford University Press.
- Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: motivation and limited resources. *Personality and Social Psychology Bulletin*, 29, 894–906.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 214–216.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3, 337–344.
- Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: more than a common problem of inhibition. *Child Development*, 73, 752–767.
- Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: you catch it from your sibs. *Child Development*, 65, 1228–1238.
- 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.
- Sabbagh, M. A., Moses, L. J., & Shiverick, S. M. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, 77, 1034–1049.
- 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.
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, 136, 241–255.
- Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: evidence for developmental trends. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 391–409). USA: Taylor & Francis Group.
- Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology*, 70, 284–288.
- Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727–741.
- Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: training children to understand belief. *Child Development*, 67, 2967–2988.
- Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, 84, 1846–1854.
- Song, H.-J., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109, 295–315.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587–592.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18, 580–586.
- Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, 30, 239–277.
- 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.
- Zaitchik, D., Iqbal, Y., & Carey, S. (2013). The effect of executive function on biological reasoning in young children: an individual differences study. *Child Development*, 85, 160–175.