The Dissertation Committee for Jessie Raye Bauer Certifies that this is the approved version of the following dissertation:

Exploring the Fundamentals of Early Causal Reasoning

Committee:

Jacqueline Woolley, Supervisor

Amy E. Booth, Co-Supervisor

Cristine Legare

Elliot Tucker-Drob

Exploring the Fundamentals of Early Causal Reasoning

by

Jessie Raye Bauer

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

May 2018

Dedication

For my mother and husband.

I would also like to thank my dog, Klaus, without whom I would have finished writing this manuscript at least a year earlier.

Acknowledgements

First, I would like to thank my supervisor, Amy Booth, for her consistent support and guidance. Over the course of three years, she has encouraged me to always be at my best, think deeply, and challenge myself. Second, I would like to thank my family for their encouragement over the years. I would like to thank my mom, who never doubted my abilities, supported me from afar, loves me unconditionally, and shaped me into the person I am today. Thank you to my husband, who has been by my side and one of my biggest champions during my entire graduate school career. Thank you to Mary Abbe, Meghan, and Heather for their friendship and love. Thank you to my committee members for thoughtful discussions and helpful suggestions. Finally, I would like to thank the hard working and dedicated research assistants and families who made this study possible.

Abstract

Exploring the Fundamentals of Early Causal Reasoning

Jessie Raye Bauer, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Jacqueline Woolley

Co-Supervisor: Amy Booth

The goal of this dissertation was to identify potential cognitive components of causal reasoning and to investigate their developmental trajectory in early childhood. We specifically focused on executive function (EF) as a potentially fundamental predictor of causal reasoning. While previous research has demonstrated that EF is related to achievement in other academic domains such as reading and math, relatively little attention has been paid to its relationship to scientific processes like causal reasoning, particularly in early childhood. To examine how EF potentially relates to the development of causal reasoning, we recruited 140 3-year-olds and 81 5-year-olds to complete three causal reasoning tasks, a battery of EF tasks, and additional cognitive measures. Results from a series of multiple regressions revealed that EF predicted contemporaneous causal reasoning, even after controlling for the influence of age, processing speed, and vocabulary knowledge. However, less variance than expected was accounted for by EF and additional covariates. We also found that a version of the traditional “blicket detector” task did not correlate with our other two measures of causal reasoning, and was not predicted by EF. Although additional research will be required to further clarify these relationships, the current results suggest that EF has the potential to support causal reasoning. Results are discussed in the broader context of scientific literacy.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The goal of this dissertation is to explore the cognitive components of early causal reasoning and to investigate their developmental trajectory in early childhood. We focus on early causal reasoning because it has the potential to serve as a critical foundation for scientific literacy. Indeed, at the very core of scientific endeavors is an understanding of causality, underlying the ability to make predictions, test hypotheses, and interpret evidence. We know that even preschoolers are able to engage in at least some forms of causal reasoning (Gopnik & Schulz, 2007). Yet little is known about individual differences in these early emerging skills, or their relationships to other cognitive processes that are undergoing rapid development during this time.

CAUSAL REASONING

Causal reasoning is the process by which we make sense of relationships between objects and events, and learn about the structure of the world around us (Gopnik et al., 2004; Kushnir, Gopnik, Schulz, & Danks, 2003). It is called upon in making inferences and deciding among competing courses of action in one's daily life. It is also essential for generating novel predictions, revising previously held beliefs, and executing successful interventions to test hypotheses in more formal contexts (e.g., Gopnik & Schulz, 2007). In this way, it is potentially foundational to scientific literacy.

A wealth of empirical work has demonstrated that young children have a natural inclination to try to understand the causal nature of objects and events (Alvarez & Booth, 2014; Gopnik, 2000; Shultz, 1982). For example, when preschoolers inquire about novel objects, they most often want to know about causally relevant properties (e.g., the functions of objects) (Greif, Kemler Nelson, Keil, & Gutierrez, 2006; Kemler Nelson, Chan Egan, & Holt, 2004). Young children will also explore novel objects more if their causal structure is ambiguous than if it is expected (Cook, Goodman, & Schulz, 2011).

But how much do young children really understand about causality? Early "generative transmission" models highlighted children's understanding of basic principles that underlie causal reasoning in the physical domain (Bullock, Gelman, & Baillargeon, 1982; Hagmayer, Sloman, Lagnado, & Waldmann, 2007; Shultz, 1982). For instance, research has shown that young children make use of the temporal priority principle (i.e., the assumption that causes must precede their effects) to interpret relationships between objects and events (Kun, 1978; McCormack & Hoerl, 2005, 2007; Shultz & Mendelson, 1975; Sophian & Huber, 1984). In one seminal experiment testing children's understanding of this principle, Bullock and Gelman (1979) showed children a large box with two ramps on each side. Children first saw a marble roll down one ramp, followed by the release of a jack-in-the-box. Then they saw another marble roll down the opposite ramp. When asked which marble made jack jump, three-year-olds were less likely than 4- and 5-year olds to make the correct causal inference (i.e., to select the first marble). Children also are sensitive to *spatial* relations between possible potential causes and their effects (e.g., Bullock et al., 1982; Leslie & Keeble, 1987). For example, in a study by Kushnir & Gopnik (2007), three-year-olds were more likely to correctly attribute causal properties to objects that were at some point spatially contiguous with a target object than to objects that never come into physical contact.

A newer model of causal learning, based on Bayesian inference, extends beyond the generative transmission model and can account for children's ability to reason about events without observable mechanisms, as well as to represent complex causal chains. Bayesian inference is a model of learning that allows children to use both prior domain specific knowledge (i.e., physical laws) and observed patterns of covariation in order to make accurate predictions and generate effective interventions (Gopnik et al., 2004; Schulz, Bonawitz, & Griffiths, 2007; Schulz & Gopnik, 2004; Sobel & Munro, 2009; Tenenbaum, Griffiths, & Kemp, 2006). This type of reasoning was demonstrated in a study by Gopnik and colleagues (2001) in which children were introduced to a "blicket detector" machine that activated (e.g., lit up or played music) when some objects (e.g., blickets), but not others, came in contact with it. As a group, children as young as three were able to correctly infer which objects were causal in accordance with the observed pattern of associations between the block placements and machine activations. In a subsequent experiment, children were able to not only identify the causal blocks, but also to infer and execute novel interventions thereon (i.e., make the machine go or stop, even when they were not explicitly shown how). Impressively, children were able to do this with only very brief observational experience. It should be noted that the anticipatory looking behavior of 8-month-olds suggests that the ability to engage in Bayesian inference begins to emerge in infancy (Sobel & Kirkham, 2006, 2007). However, when children must make novel interventions (i.e., make the machine go or stop) or make inferences based on more complicated rules, a significant shift in performance is observed from 3 to 5 years of age (e.g., Fernbach, Macris, & Sobel, 2012; Sobel, Tenenbaum, & Gopnik, 2004).

This later developmental shift is also evident in tasks that challenge children to draw upon existing knowledge to reason about more naturalistic causal sequences. For instance, Gelman, Bullock & Meck (1980) showed 3- and 4-year-olds pictures of events in which objects underwent changes (e.g., an intact cup next to the same cup but in a broken state). Children at both ages were able to choose the correct causal instrument (e.g., hammer) from an array of pictured alternatives. However, other work has found that 4-year-olds were more successful than 3-year-olds in making inferences about complex state changes involving more than one dimension of transformation and/or atypically reversed causal transformations (e.g., a dirty shirt changing to a clean shirt) (Das Gupta & Bryant, 1989; Gelman, Bullock, & Meck, 1980). For example, in a study by Hong and colleagues (2005), they found a significant developmental shift in children's performance on a causal reasoning task wherein only 10% of 3-year olds passed, compared to 52% of 4.5-year olds. Other studies have demonstrated similar developmental results, especially as task complexity increases (Frye, Zelazo, Brooks, & Samuels, 1996; Frye, Zelazo, & Palfai, 1995)

Perhaps not surprisingly, even more protracted developmental changes have been observed in more demanding measures of causal reasoning that involve explicit consideration of counterfactuals (i.e., situations that are inconsistent with reality). In one relevant study, Guajardo & Turley-Ames (2004) told children a story about a child playing in a muddy backyard. The child in the story becomes thirsty, and then goes inside to get some juice. Consequently, the floor inside gets muddy. When asked what the child in the story could have done so that the floor would *not* have gotten dirty, four and five-year-old children performed significantly better than three-year-olds in choosing among pre-formulated counterfactual solutions offered by the experimenter (e.g., they could have taken off their shoes before entering the house) (Guajardo & Turley-Ames, 2004; see also, Drayton, Turley-Ames & Guajardo, 2011). In order to provide a correct response on this task, young children needed to engage in a highly complex reasoning process in which they held certain events constant, while also manipulating others to determine causally relevant features. When, in subsequent work, children were required to produce, rather than simply select, appropriate counterfactuals (e.g., Harris, German & Mills, 1996) children struggled until 6 years of age (Beck, Burns, & Riggs, 2011; Rafetseder & Perner, 2014; Rafetseder, Renate, & Perner, 2010; Rafetseder, Schwitalla, & Perner, 2013). Other work has suggested that slower development might be particularly evident when children are asked to reason counterfactually about positive outcomes (e.g., she was not hungry at the end of her trip because she chose to eat a sandwich instead of only eating a piece of candy) (German, 1999).

Taken together, these findings suggest that causal reasoning skills are present at an early age and undergo significant development throughout the preschool years. While three-year-old children were in general able to track objects and associated events, and use patterns of covariation to make causal inferences at above chance levels, they often failed to perform optimally. By five years of age, children made errors on causal reasoning tasks far less frequently. Indeed, older children appeared able to go beyond adhering to simple causal principles in order to incorporate new evidence, revise previously held beliefs if necessary, and engage in complex reasoning.

Despite the early development of causal reasoning and its potential for shaping long-term scientific literacy and academic achievement, we have little understanding of its structural origins. Specifically, is causal reasoning a unique capability that develops relatively independently or is it largely dependent on broader, more fundamental, cognitive abilities? Although we consider a number of potential contributors to the development of causal reasoning in this project, we will be particularly interested in executive function (EF). EF is an especially promising contributor based on its already established role in the development of core academic domains like reading and math (Best, Miller, & Naglieri, 2011; Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; St Clair-Thompson & Gathercole, 2006). Although its relevance to understanding processes in the scientific domain has received relatively little attention (Gropen, Clark-Chiarelli, Hoisington, & Ehrlich, 2011; Zaitchik, Iqbal, & Carey, 2014), EF has been associated with a broad array of related cognitive skills such as planning and goal directed problem solving (Best et al., 2011; Blair & Razza, 2007; Brock et al., 2009; Diamond, 2013).

EXECUTIVE FUNCTION

EF refers to a set of skills that control and regulate attention, action, and behavior (e.g., Best et al., 2011; Huizinga, Dolan, & van der Molen, 2006; Lee, Bull, & Ho, 2013; Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003). For the purposes of this dissertation, we follow Miyake et al (2000) and conceptualize EF as an integrative framework with partially dissociable factors (see also, (Garon, Bryson, & Smith, 2008). We focus on three well-recognized components of EF: working memory, inhibition, and cognitive flexibility.

Here we define working memory as the ability to hold and manipulate information in mind (Baddeley & Hitch, 1994). The fundamental ability to hold information in mind emerges early. Indeed, evidence suggests that even infants can remember the location of objects hidden in the context of A-not-B type tasks (Diamond 1985; Diamond 1995, Nelson et al. 2012). However, the ability to mentally manipulate that active information develops slowly over early childhood (e.g., Cowan et al. 2002, 2011; Crone et al. 2006). Dual tasks, or tasks that require alternating between two activities that are contingent upon each other, are a common approach used to test working memory ability. One frequently used example involves asking adults to remember an increasing number of digits while answering comprehension questions (Baddeley & Hitch, 1974). In order to adapt these types of tasks for use with children, researchers have incorporated friendly puppets that use simple language to impart instructions. In the count and label task (Gordon & Olson, 1998), for instance, a puppet asked children to alternate between counting and naming familiar objects in front of them. A large study by Carlson (2005) showed that count and label performance improved with age, with 26, 55, 71, and 77% of 3-, younger 4-, older 4-, and 5-year olds passing this task, respectively.

The second component of EF, inhibition, involves effortful control of prepotent behaviors, thoughts and attention patterns (Diamond, 1990). Inhibition has a far more protracted developmental timeline than working memory (Diamond, 2013). Indeed, while limited inhibition skills emerge in early childhood, these undergo considerable subsequent development, even into adolescence. Broadly speaking, inhibition includes two distinct subcomponents: attentional and behavioral inhibition. Attentional inhibition requires selective attention, or filtering out irrelevant or distracting information. A common measure of attention inhibition is the Flanker task (e.g., Eriksen & Eriksen 1974) wherein the participant is required to attend and respond to a central stimulus, while ignoring surrounding stimuli. The NIH-Toolbox (NIH-TB) includes a version of the flanker task for young children, in which they must indicate the direction that a central fish is swimming, while ignoring the directions of the other surrounding fish. Stroop tasks are another common method used to assess attentional inhibition. The black and white stroop task, for instance, requires that children inhibit the preponent response of naming the color on the card (e.g., saying “white” when they are shown a white card), while naming the opposite color (i.e., “black”). In contrast, behavioral inhibition is typically tested by setting up situations in which children must resist producing a prepotent action, like in a game of Simon Says (e.g., Hommel, 2011). For example, in the bear/dragon task (Reed, Pien, & Rothbart, 1984), which is adapted for particularly young children, the goal is to imitate the actions of a nice bear puppet, but not those of a mean dragon puppet. This task reveals significant developmental differences in preschoolers wherein young 3-year-olds perform at chance, but reliably succeed by age five (Carlson, 2005).

Finally, the third component of EF, cognitive flexibility, is the ability to adjust to new tasks and demands, or to change perspectives (Diamond, 2013; Kushnir et al., 2003; Zelazo & Frye, 1997). Although cognitive flexibility can be observed in preschool populations, like other aspects of EF, it undergoes considerable development throughout early childhood. Cognitive flexibility in children is often measured with the Dimensional Change Card Sort Task (DCCS; Zelazo et al., 1996). In this task, children are asked to sort stimuli according to a rule (e.g., by color). Then, after children demonstrate mastery of the first rule, they are instructed to use a different rule (e.g., by shape). Children younger than four-and-a-half often perseverate (e.g., continue to sort by color) and have difficulty switching to sorting by the new rule.

Although these three aspects of EF have been studied independently, it is important to note that they are highly interconnected and often difficult to disentangle, especially in young children. Indeed, some have argued that some of the most widely used measures of EF tap multiple components thereof. For example, the DCCS taps both working memory (by requiring that one hierarchical action rule be held in mind) and inhibitory control (by requiring that stimulus dimensions irrelevant to that rule be ignored), and thus is not purely a measure of cognitive flexibility (Carlson, 2005; Diamond, 2005; Doebel & Zelazo, 2015). Diamond (2013) goes further to argue that inhibition and working memory are fundamental precursors to cognitive flexibility. In studies aimed at formally describing the conceptual structure of EF in children, researchers administered a large battery of EF tasks designed to tap all three potential dimensions (working memory, inhibition, and flexibility) to varying degrees. Factor analyses with preschoolers have yielded mixed results, though models where all the tasks load onto one common unifying factor have received the most support (Diamond, 2013; Hughes, 1998; Hughes & Ensor, 2011; M. R. Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012; Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011). It is not until middle childhood, around 8-11 years of age, that the three factors comprising EF appear to become clearly differentiated (e.g., Lehto et al., 2003).

A helpful framework for understanding the development of EF is the interactive specialization model (ISM) (Johnson, 2011). According to the ISM, brain regions (and corresponding EF skills) are not initially finely tuned or optimally integrated. Instead, over time, brain regions associated with EF gradually become interconnected, allowing for quantitative developments in EF. Evidence suggests that rapid periods of growth in the prefrontal cortex (and its connections to other brain regions) are particularly relevant to understanding the development of EF (see Blair & Razza, 2007; Carlson, 2005; Diamond, 2002; Diamond, Carlson, & Beck, 2005; Müller, Zelazo, Hood, Leone, & Rohrer, 2004; Wiebe et al., 2011; Zelazo, 2006). Preschool represents one such period of rapid growth. As already reviewed, a wealth of evidence demonstrates that EF, as a whole, undergoes substantial quantitative development in efficiency during the preschool years (Carlson & Moses, 2001; Espy, Kaufmann, McDiarmid, & Glisky, 1999; Kirkham, Cruess, & Diamond, 2003). In addition, preschool has been described as a period of qualitative transition, marking a shift from reflexive (i.e., impulse driven) to reflective (i.e., pausing to consider multiple choices) thoughts and behaviors (Carlson & Moses, 2001). In other words, children begin to have some control over regulating their emotions and behaviors, and develop the ability to think through far-ranging possibilities before reaching a conclusion or initiating a response. This purported shift from reflexive to reflective thought is also consistent with Bayesian models in which preschool children develop the ability to consider and think through multiple possible predictions, and then select the best choice (Gopnik & Glymour, 2002; Gopnik et al., 2004).

JOINT DEVELOPMENT OF EXECUTIVE FUNCTION AND CAUSAL REASONING

Gropen, Clark-Chiarelli, Hoisington, & Ehrlich (2011) provide a theoretical framework that helps clarify how EF might support causal reasoning. The authors suggest that EF scaffolds a child's ability to engage in the simultaneous representation of multiple nested rules and conditionals necessary for processing cause and effect relations in complex problem spaces (Anderson, 2002). More specifically, working memory might be important for causal reasoning because it allows us to bring together incoming perceptual information and past knowledge to resolve, and think creatively about, novel causal systems (Byrne, 2007). At the same time, inhibition might play a critical supporting role by helping to focus attention on relevant dimensions of the causal system and by suppressing, and allowing the revision of, prior beliefs (Gopnik, Sobel, Schulz, & Glymour, 2001; Legare, Gelman, & Wellman, 2010; Shultz, 1982; Wilkening & Sodian, 2005). Empirical work has provided evidence for relationships between both the working memory and inhibition components of EF and general reasoning abilities in adults and older children. For example, Handley et al. (2004) found that both working memory and inhibition positively predicted 10- to 11-year-olds' performance on transitive and conditional reasoning tasks. Other work has shown that working memory and inhibition positively predicted performance on reasoning ability with premises that are empirically false (i.e., not dependent on prior knowledge or experience) in both children (Simoneau & Markovits, 2003) and young adults (Markovits & Doyon, 2004). While this literature does not address *causal* reasoning specifically, it is consistent with the idea that EF might be involved in advanced reasoning capabilities generally speaking.

Work with younger children has more directly addressed relationships between EF and causal reasoning, specifically as instantiated in counterfactual reasoning tasks. For example, Guajardo, Parker & Turley Ames (2009) found that both representational flexibility and working memory correlated with 3, 4-, and 5-year olds' counterfactual reasoning ability. Beck et al. (2009) failed to replicate this effect in 3- and 4-year-olds, but did report a significant relationship between inhibition and counterfactual reasoning performance in this age group (Beck et al., 2009; Kevin J Riggs & Beck, 2007). Although some have argued that these discrepancies indicate independence of counterfactual reasoning from the development of EF (Rafetseder et al., 2013), the weight of the evidence suggests to us that EF likely contributes to causal reasoning from its inception.

By examining some of the most prominent measures of causal reasoning used with children, it might be possible to isolate potential points of EFs' influence. For example, working memory might be necessary for keeping track of block placements and machine activations in "blicket detector" tasks (Gopnik, 2000). It might also be important for integrating that observed evidence with prior knowledge (e.g., about the importance of contingency in determining causes). Inhibition might also be required for suppressing compelling alternatives (i.e., non-blickets) in the process of inferring which object caused the observed sequence of activations. Similarly, when faced with the forced choice decisions typical of causal inference tasks, inhibition might be essential to ignoring the incorrect lures. Cognitive flexibility might also be key when reasoning about reverse transformations that are misaligned with prototypic experiences (and, therefore, expectations). And working memory might be particularly important for reasoning about more complex state changes. For example, inferring the cause of an intact cup changing into a broken *and* wet cup will require more working memory than inferring the cause of an intact cup more simply changing into a broken cup. Indeed, Hong et al. (2005) found that children's performance declined as the complexity of causal relations increased (e.g., from if-then to if-if-then), especially for younger children. EF also likely contributes to the ability to reason counterfactually about causal scenarios. For example, working memory might play a role in generating plausible alternative scenarios or explanations while also holding what actually happened in mind (Robinson & Beck, 2000). Inhibitory control and cognitive flexibility might be particularly important in temporarily ignoring what is known to be true in order to consider alternatives (Bonawitz et al., 2011; Cook et al., 2011; Legare, 2012; Marcis & Sobel, 2017). In sum, both causal reasoning and EF undergo significant developments between three and five years of age (e.g., Gopnik & Sobel, 2000; Hong et al., 2005), and both evidence and theory suggest that these skills might be related in preschoolers, older children, and adults.

THE CURRENT STUDY

Broadly construed, the goal of the current study is to begin specifying candidate foundations of causal reasoning in children. In pursuit of this goal, we will consider whether causal reasoning develops independently as a unique cognitive capacity or whether it develops in tandem with domain-general processing skills. More specifically, this study will focus on EF as an especially promising candidate for supporting the development of causal reasoning. Indeed, the studies reviewed above all demonstrate some relationship of EF to causal (or broader) reasoning ability despite the varied methodological differences (e.g., varying numbers and types of EF measures) and different age groups tested. We therefore hypothesize that the development of causal reasoning will rest on parallel developments in EF. Although our study design cannot directly test this contingency, we can evaluate its plausibility by examining correlations between measures of EF and causal reasoning across a period of rapid developmental change. We will evaluate patterns of covariation using structural equation modeling (SEM) and regression approaches. This investigation is important because, although previous work has begun to explore relationships between EF and causal reasoning at different developmental time points, the evidence remains inconclusive. The current work will help to clarify inconsistencies in the literature and specify the potential for EF to support the development of causal reasoning.

Chapter 2: Methods

PARTICIPANTS

Data for this study was collected as part of a larger longitudinal study investigating the development of children's interest in causal information. Our study sample included 221 children (81 = 5-year-olds, 140 = 3-year-olds) from the Austin, Texas area. Participating children were three or five years of age at the first session (M_3 = 40.89 mo; SD_3 = 3.15 mo, range = 36.14-47.27 mo, 77= female; M_5 = 65.19 mo; SD_5 = 3.79 mo, range = 56.24-71.98 mo, 45 = female). Children did not have any diagnosed developmental delays or disorders, and they understood English "well" or "very well" as reported by a parent.

The sample was racially, ethnically, and socioeconomically representative of our recruitment area. Based on parent report, 12.67% of participating children were African American, 68.78% were White, 4.07% were Asian or Pacific Islander, and 14.48% were classified as multiple races or other. About 32.13% of these children were also identified by their parents as being Hispanic or Latino.

With respect to maternal education, 0.9% of mothers reported not completing high school, 18.55% held a high school degree, 12.67% completed some college or additional training beyond high school, 38.91% had a four-year bachelor's degree, 26.24% held a master's degree or higher, and 2.71% declined to report their level of education.

PROCEDURE

Data for this study were collected over two to three sessions lasting approximately 45-60 minutes each. Sessions were audio-visually recorded for offline coding of participant responses and verification of protocol fidelity. The first session for three-year-olds took place at a local children's museum. The remaining sessions took place in our laboratory in a colorfully decorated room at a child-sized table. At the first session, parent consent was obtained and children were assessed on receptive language (Picture Vocabulary Test). Parents also completed a demographic interview as part of the larger longitudinal study that included questions concerning parental education, race and ethnicity, as well as household composition, income, and literacy environment. Children completed EF and causal reasoning measures during the two subsequent laboratory sessions. In the first laboratory session, children completed the cause-effect association, black/white stroop, flanker, causal inference, digit span, and count/label tasks. At the second laboratory session, children completed the counterfactual reasoning task, DCCS, bear/dragon, and Processing Speed tasks. The testing sessions for the 5-year-olds were identical, except the receptive language measure was also administered at the first laboratory session.

Measures

Measuring Executive Function

We used several measures of EF tapping working memory, inhibitory control, and cognitive flexibility.

Working Memory

The count and label task was used to assess working memory, or the ability to hold and manipulate information in one's mind (Baddeley & Hitch, 1994; Gordon & Olson, 1998). We chose this particular test because it is one of few options available that is developmentally appropriate for very young children (Carlson, 2005). In this task, children were shown three objects (e.g., a key, a cup, and a toy dog) and asked to label them. Then the experimenter suggested that they count the objects (e.g., "one", "two", "three"). Then they demonstrated how to count the objects and label them each in turn (e.g., "one is a key, two is a cup, three is a dog") and asked the child to try. Children were assigned a score of either 0 (for incorrect responses) or a score of 1 (for correct responses) on each of two trials. This task took approximately 2 minutes to administer. The outcome variable is a raw score ranging from 0-2.

Inhibitory Control

The NIH-TB Flanker Inhibitory Control and Attention Test (Flanker; Zelazo et al., 2013) was used to evaluate children's inhibitory control of visual attention (Gershon et al., 2013). During each trial of this task, which was presented on an iPad, a central target fish was flanked by similar stimuli on the left and right. Children were instructed to indicate the direction of the central stimulus by pressing a left or right arrow button on the screen. On congruent trials, the flanker fish faced the same direction as the target fish. On incongruent trials, they faced the opposite direction, requiring the child to inhibit their attention to the flanker fish. Total administration time for the adaptive test is approximately 5 minutes. The flanker task was normed on a sample of 174 children. Zelazo and colleagues (2013) reported the NIH-TB Flanker positively correlated with the WPPSI–III Block Design in 3 to 6 year olds ($r = .60$) and D-KEFS Inhibition raw scores in 8 to 15 year olds ($r = .34$). The outcome variable reported is an uncorrected standard score based on either accuracy or, if a child achieved greater than 80% accuracy, a combination of accuracy and reaction time (normative mean = 100, SD = 15).

In the black and white stroop task (Vendetti, Kamawar, Podjarny, & Astle, 2015), children were shown a white card and a black card. Children were instructed that in this game they are to say "black" for the white card and "white" for the black card. After a brief warm-up, there were 21 test trials with each card presented (from beneath the table) in a fixed, pseudorandom order and shown for one second. There were no breaks or rule reminders and self-corrected answers were not counted. If children did not respond within the one-second window, the experimenter prompted them to make their best guess. If children could not remember or were not sure, we did not score the trial. This task took approximately 2 minutes to administer. Therefore, the dependent variable is proportion trials correct, with a continuous range from 0-1.

In the bear/dragon task (Reed et al., 1984), the experimenter introduced children to a "nice" bear puppet and a "naughty" dragon puppet. Children were told they are to do what the bear asks them to do (e.g., "touch your nose!") but not to do what the dragon asks. After practicing, there were 10 test trials with the bear and dragon commands in alternating order. All actions involved hand movements. The variable of interest in this task is an index of self-control measured by the number of dragon trials where the child succeeded in inhibiting an action. This task also took approximately 2 minutes to administer. Children received a score of either a 0 (for movement) or a 1 (no movement) for each dragon trial for a possible range of 0 to 5.

Cognitive Flexibility

The NIH-TB Dimensional Change Card Sort Test (DCCS; Zelazo et al., 2013) was used to evaluate cognitive flexibility, or the ability to adjust to new tasks and demands (Bullock et al., 1982). On each trial of the NIH-TB DCCS, a target visual stimulus must be matched to one of two stimuli according to shape or color. Children first received a block of trials in which one dimension (e.g., shape) was relevant to this decision, and then a second block (switch) in which the other dimension (e.g., color) was critical. Those who succeeded following the switch also received a mixed block, in which shape was relevant on most trials, with occasional and unpredictable shifts to color. The relevant criterion word, "color" or "shape," was simultaneously presented in visual and auditory form. Total administration time for the adaptive test is approximately 5 minutes. This task was normed on a sample of 166 children, and Zelazo et al. (2013) reported significant correlations with the WPPSI–III Block Design in 3 to 6 year olds (r = .69), and D-KEFS Inhibition raw scores in 8 to 15 year olds (r = .64). Similar to the NIH-TB Flanker task, if children achieve 80% accuracy or higher, the algorithm weights accuracy and reaction time in the uncorrected standard score (normative mean = 100, SD = 15).

Measuring Causal Reasoning

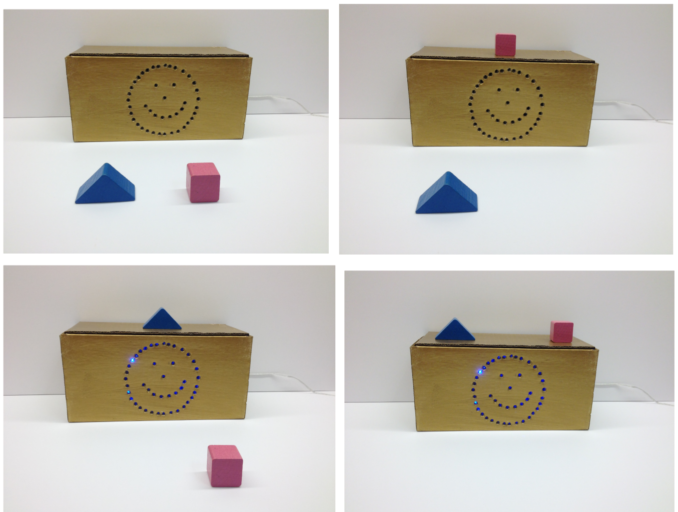
Causal reasoning was assessed with three measures that tapped cause-effect association, causal inference, or counterfactual reasoning.

Tracking Cause-Effect Associations

Four boxes ("blicket detectors") measuring 8 x 4 x 4 inches were constructed out of cardboard (Gopnik & Sobel, 2000; Gopnik et al., 2001) (see Figure 2). The first box was painted blue and had a gear toy affixed to the top that spun when the box was activated. The second box, painted yellow with blue polka dots, produced a continuous playful noise when activated. The third box was painted green and had a laminated yellow cartoon cat figure mounted on the front. When activated, a small light bulb inside the box illuminated, making the cat's glow. The fourth box was painted silver and had a puppy cartoon on the front. When activated, a motor within the box slowly turned the puppy back and forth.

The experimenter introduced the task to the children by saying, "I have some special boxes with me! Some blocks make the boxes go, and some blocks don't do anything to the boxes. They don't turn them on at all! Do you think you can help me figure out how the boxes work?" A single block was then introduced with a unique novel name and was placed on the box to no effect (e.g., "This one is a duffin. This one doesn't do anything to our boxes."). The first block ("duffin") was then removed and the second block (e.g., a "buttle") was introduced and then placed on the box, which led to its activation. Finally, while the box was still activated, the experimenter also placed the non-causal block (the "duffin") back on the box while simultaneously saying, "let's put both on together!" The box remained active while both blocks were on the box. The experimenter then asked the child, "Now I need your help! How can we make the box stop?" In order to respond correctly, children had to keep track of which object was more reliably associated with activation of the box and remove it. If a response was not immediately forthcoming, children were further prompted (e.g., "Can you tell me which one?" or "Which one do you think?") until they offered an answer. This same procedure was repeated for the remaining three trials. This task took approximately 5-7 minutes to administer. Children were assigned a score of 1 if they chose the correct block and a score of 0 if they chose the incorrect block on each of the four trials. Although the experimenter discouraged children from choosing both blocks by holding both blocks down with their hands and asking the child to choose just one, children did not always comply, therefore leading to the occasional uncodable trial. In these rare circumstances, we excluded the trial from analysis. Therefore, the outcome variable is proportion trials correct with scores ranging from 0-1.

Figure 1. Example Trial from the Cause-Effect Association Task.



Note. Panel 1: "This is a special machine! Panel 2: "Here is one that doesn't make it go, it's called a mab." Panel 3: "Here is one that makes it go, it's called a veep." Panel 4: "Let's put both on and watch it again! Now I need your help! How can we make the box stop?"

Generating Causal Inferences

This task, modeled after previous studies (Das Gupta & Bryant, 1989) involved eight trials (see Table 1). On each trial, children were shown a timeline of photographs in which objects underwent changes (e.g., a broken flowerpot next to a whole flowerpot). The experimenter first described the pictures in general terms ("First it looked like *this*, then I did something to it, and now it looks like *this*."). Then, children were shown photographs of four possible instruments (e.g., hammer, light bulb, paintbrush, glue) and were asked to choose the one that caused the change. This task took approximately 5 minutes to administer. The outcome variable is the total number of trials answered correctly out of eight.

Table 1. Stimuli Used In The Causal Inference Task.

Stimuli	Choices
Practice Items	
Tomato → Sliced Tomato	whisk, spatula, measuring cups, knife
Sweater with Hole → Sewn Sweater	teapot, roller-skate, mug, needle and thread
Test Items (Forwards)	
Spilled Dirt → Swept Dirt	chair, tissue box, clock, broom
Raw Egg → Cooked Egg	blender, drying rack, napkin holder, stove
Messy Hair → Brushed Hair	sponge, toothbrush, rolling pin, hairbrush
Torn Paper → Taped Paper	keys, toy blocks, crayons, tape
Test Items (Backwards)	
Wet Pan → Dry Pan	sink, microwave, calculator, tape
Chalkboard with Writing → Erased Chalkboard	chalk, scissors, stapler, eraser
Broken Flowerpot → Glued Flowerpot	hammer, light bulb, paintbrush, glue
Stained Shirt → Clean Shirt	ketchup, purse, iron, detergent

Note. Based on Das Gupta and Bryant (1989).

Figure 2. Example Trial from the Causal Inference Task.

Counterfactual Reasoning

In this task, modeled after Guajardo & Turley-Ames (2004), children were presented with four vignettes and asked to reason counterfactually about each situation. For example, in one story, children were asked to imagine that they are playing outside in a muddy yard. Then they imagine that they get really thirsty, so they go inside their house for a drink and dirty the kitchen floor in the process. Children were then asked, “What could you have done so the floor would *not* have gotten dirty?" In order to respond correctly, children had to mentally represent what would have had to have happened in the past to produce a different outcome. This task took approximately 5 minutes to administer. Children were assigned a score of 1 if they generated a counterfactual statement (e.g., take off my shoes) on each of four trials. They were otherwise assigned a score 0 for each trial. The outcome variable is therefore a score with a range of 0-4.

Figure 3. Example Trial from the Counterfactual Reasoning Task.

Note. Panel 1: "Imagine that you are playing outside in the muddy yard." Panel 2: "You are thirsty so you go inside to the kitchen to get a drink of juice." Panel 3: "You walk through the mud, you step over the doormat, and you keep your shoes on. Because your shoes are muddy you get dirt all over the floor!" Test question: "What could you have done so the kitchen floor would not have gotten dirty?"

Measuring Other Cognitive Factors

Two different measures from the NIH Toolbox Cognition Battery (Bauer & Zelazo, 2014; Zelazo et al., 2013) were used to measure different aspects of cognition.

Receptive Vocabulary

The NIH-TB Picture Vocabulary Test (PVT; Gershon et al., 2013) is a measure of receptive vocabulary administered in a computerized adaptive format on an iPad. During this task, children were presented with an audio recording of a word while four pictures simultaneously appeared on the screen. On each trial, children were asked to select the picture that most closely matched the meaning of the word that was said. The child was presented with two practice trials with feedback, followed by 25 test trials. Total administration time for the adaptive test is approximately 5 minutes. Initial item-level calibration was conducted online with 3,190 children ages 3 to 17 (Gershon et al., 2013). The version of the PVT used in this study was normed on a sample of 120 3-6-year-old children and correlates strongly (r = .9, p < .001) with the Peabody Picture Vocabulary Test 4th Edition (Gershon et al., 2013; Weintraub et al., 2013). As is the case for the other NIH-TB tasks, the NIH-PVT also yields an uncorrected standard score (normative mean = 100, SD = 15).

Processing Speed

On each trial of the NIH-TB Pattern Comparison Processing Speed Test (Processing Speed; Carlozzi, Tulsky, Kail, & Beaumont, 2013)) requires participants to discern whether two side-by-side pictures are the same or not. Younger children make this decision by choosing either a “smiley face” (corresponding to a “yes” response) or a “frowny face” (corresponding to a “no” response). In this task, the dependent variable is an uncorrected standard score based on the raw number of trials answered correctly in a 90s period (normative mean = 100, SD = 15).

CODING

Trained researchers scored the cause-effect association, causal inference, counterfactual reasoning, count and label, black and white stroop, and bear/dragon tasks offline using video recordings. For all tasks, a second coder assessed 20% of files to ensure reliability of coding. The coders also assessed videos for fidelity to protocol on the bases of general procedures and adherence to a script (if applicable). In addition, for the counterfactual reasoning task, each video was initially transcribed and coded by a blind researcher, then a second researcher coded the transcript. There was excellent agreement for the dichotomous coding scheme used for the counterfactual reasoning task (Cohen's $\kappa = .93$ and $.95$, for 3- and 5-year-olds, respectively). All four measures from the NIH-TB were automatically scored. Participant data were coded and managed using REDCap (Research Electronic Data Capture) hosted at University of Texas at Austin (Harris et al., 2009). REDCap is a secure, web-based application designed to support data capture for research studies.

Chapter 3: Results

MISSING DATA

Rates of missing data range from 0% to 16.28% (see Table 2) across tasks.

Three-Year-Olds

We excluded 28 3-year-olds who did not meet the age criterion (*n* = 1), poor overall behavior (*n* = 7), not meeting the English language requirement. We also removed participants with less than half of the causal reasoning measures (*n* = 5), EF measures (*n* = 13), or both (*n* = 2). Additional task level data was excluded on a case-by-case basis because children received an attention/behavior score of a 1 or 2 (Processing Speed: *n* = 3; count/label: *n* = 6; flanker: *n* = 3; stroop: *n* = 3; bear/dragon: *n* = 2; DCCS; *n* = 3; causal inference: *n* = 3; counterfactual reasoning: *n* = 2), and failure to pass training trials (Processing Speed: *n* = 3; flanker: *n* = 3; stroop: *n* = 3; DCCS: *n* = 8). We lost additional task data due to experimenter error (NIH-Processing Speed: *n* = 7; count/label: *n* = 4; stroop: *n* = 1; bear/dragon: *n* = 4; counterfactual reasoning: *n* = 2), and attrition (Processing Speed: *n* = 4; flanker: *n* = 2; stroop: *n* = 3; bear/dragon: *n* = 6; DCCS: *n* = 5; causal inference: *n* = 2; counterfactual reasoning: *n* = 5).

Five-Year-Olds

We excluded 15 5-year-olds with a clinical diagnosis (*n* = 2), not meeting the age criterion (*n* = 1), and failure to return to the second session (*n* = 12). We also removed participants with less than half of the causal reasoning measures (*n* = 2). Additional task level data was excluded on a case-by-case basis because children received an attention/behavior score of a 1 or 2 (CEA: *n* = 1), and failure to pass training trials (flanker: *n* = 1; DCCS: *n* = 1). We lost additional task data due to experimenter error (CEA: *n* = 1; count/label: *n* = 1; stroop: *n* = 1; counterfactual reasoning: *n* = 8; flanker: *n* = 2; bear/dragon: *n* = 3; PVT: *n* = 6).

Table 2. Rates of Missingness for Each Study Variable for Final Sample.

	% Missing	
Variable	3-year-olds	5-year-olds
Vocabulary	0	9.23
Processing Speed	16.28	4.62
Count/Label	12.07	1.54
Flanker	6.90	4.62
Stroop	11.21	1.54
Bear/Dragon	12.07	4.62
DCCS	15.52	1.54
CEA	0.86	6.15
CI	6.90	0
CFR	8.62	12.31

Note. DCCS = Dimensional Change Card Sort; CI = causal inference; CFR = counterfactual reasoning; CEA = cause-effect association.

DATA ANALYSIS PLAN

Missing Data

To handle missing data in our regression analyses, we employed multiple imputation (MI) using the "mice" package that also runs in RStudio (Brand, 1999; van Buuren & Groothuis-Oudshoorn, 2011). MI (Brand, 1999; Rubin & Schenker, 1986;

Schafer, 1997) allowed us to retain as many subjects as possible by replacing missing values with unbiased estimates based on the observed data (Little & Rubin, 2002). We imputed 9 and 5 datasets for 3- and 5-year olds, respectively based on the recommendation to impute the number of datasets equal to the average percent of missing data (Bodner, 2008; Royston & White, 2011; White, Royston, & Wood, 2011). We also imputed using 30 iterations, based on recommended procedures from simulation work (Brand, 1999; Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006; van Buuren & Groothuis-Oudshoorn, 2011)

Treatment of Variables

First, we scaled individual EF and causal reasoning measures (*M* = 0, range: -3, 3) for each age group separately. Then, we created a composite EF variable from all of the EF tasks including count/label, flanker, stroop, bear/dragon, and DCCS tasks. Based on a priori hypotheses and patterns of bivariate correlation observed in the data, our dependent causal reasoning variable for all analyses is a composite of causal inference and counterfactual reasoning. The cause-effect association task was analyzed separately due to its failure to correlate with either of the other causal reasoning measures. We controlled for other potential contributors to causal reasoning including child age, vocabulary, and processing speed by adding them into the models as covariates.

Data Analysis

All data analyses were conducted using RStudio (R version 1.0.136; RStudio Team, 2016; R Core Team, 2016). Recall that our core goal was to evaluate whether causal reasoning is related to EF, and to specify how this relationship changes over time. Note that we initially intended to use structural equation modeling (SEM) techniques to model the structure of causal reasoning and EF relative to one another. However, because our final sample size was smaller than originally planned (especially for the 5-year-olds), we opted for a multiple regression approach, which was more amenable to our current sample size. For the interested reader, results from the originally planned, but underpowered, SEM approach can be found in the Appendix. Because we were interested in exploring potential developmental changes in the relation between EF and causal reasoning, we present results from the 3- and 5-year old samples separately.

DESCRIPTIVE STATISTICS

The means, standard deviations, and ranges for age, vocabulary, EF measures, and causal reasoning measures are displayed in Tables 3 and 4. Simple bivariate correlations between these variables of interest are presented in Tables 5 and 6. Please recall that, because children occasionally did not provide a clear response on a single trial during the cause-effect association and stroop tasks, we use proportion correct (ranging from 0-1) as our dependent variable for these tasks. Because sex did not correlate with any of the tasks, it is not included, and was not considered in further analyses.

Table 3. Descriptive Statistics for Three-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
Age	40.97	3.17	36.17	47.27
Vocabulary	57.26	6.37	37	88
Processing Speed	48.29	8.42	31	74
Count/Label	0.79	0.90	0	2
Flanker	38.75	12.93	22	87
Stroop[a]	0.44	0.33	0	1.00
Bear/Dragon	3.66	1.93	0	5
DCCS	47.80	13.53	34	84
CEA[a]	0.64	0.32	0	1
CI	4.71	1.84	0	8
CFR	1.35	1.32	0	4

Note. DCCS = Dimensional Change Card Sort; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning. [a]Proportion correct.

Table 4. Descriptive Statistics for Five-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
Age	65.21	3.56	60.22	71.98
Vocabulary	68.09	7.71	52	91
Processing Speed	64.57	10.88	43	88
Count/Label	1.62	0.70	0	2
Flanker	68.80	17.26	24	92
Stroop[a]	0.66	0.29	0	1
Bear/Dragon	4.72	0.78	0	5
DCCS	71.60	16.21	38	96
CEA[a]	0.90	0.19	0.25	1
CI	6.98	1.19	2	8
CFR	2.80	1.33	0	4

Note. DCCS = Dimensional Change Card Sort; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning. [a]Proportion correct.

Three-Year-Olds

Although some measures of EF were positively related to each other, several correlations were small or non-significant (See Table 5). Indeed, the only significant correlations observed among EF measures were between the flanker task and both the count and label and DCCS tasks. With respect to the causal reasoning measures, counterfactual reasoning and causal inference emerged as significant correlates of each other. Correlations between these causal reasoning measures and EF measures were also evident, albeit more consistently for causal inference than for counterfactual reasoning. Specifically, causal inference correlated significantly with flanker, stroop and bear/dragon, while counterfactual reasoning correlated only with count/label. Somewhat surprisingly, performance on the cause-effect association (i.e., blicket) task failed to correlate with any key measures of EF or causal reasoning, with the exception of the DCCS. Note that both children's receptive vocabulary and processing speed correlated with measures of causal reasoning, thereby necessitating that we control for their influence in our analyses.

Table 5. Correlations Between Main Study Variables for Three-Year-Olds After Imputation.

	Age	Vocab.	Speed	Count	Flanker	Stroop	Bear	DCCS	CEA	CI
Age										
Vocab.	0.25**									
Speed	0.26*	0.25**								
Count	0.22*	0.17†	0.44***							
Flanker	0.25**	0.23*	0.07	0.3**						
Stroop	0.13	-0.16†	0.09	0.19†	0.2†					
Bear	0.18†	0.22*	0.27*	0.21†	0.06	0.05				
DCCS	0.16	0.04	-0.14	0.08	0.23*	0.09	0.08			
CEA	0.1	-0.04	-0.27*	-0.03	0.11	-0.01	-0.06	0.22*		
CI	0.3**	0.41***	0.15	0.18†	0.25**	0.23*	0.19*	0.07	0.04	
CFR	0.35**	0.18†	0.23*	0.27**	0.15	0.07	0.19†	0.08	-0.16†	0.26**

Note. Simple correlation coefficients are shown; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; Speed = processing speed; †*p*<0.1; **p*<0.05; ***p*<0.01; ****p*<0.001.

Five-Year-Olds

Patterns of correlations were similar for the 5-year-olds, with a few exceptions (see Table 6). First, the DCCS now correlated with flanker, in addition to count and label. Second, counterfactual reasoning now correlated with flanker and bear/dragon, in addition to count/label.

Table 6. Correlations Between Main Study Variables for Five-Year-Olds After Imputation.

	Age	Vocab.	Speed	Count	Flanker	Stroop	Bear	DCCS	CEA	CI
Age										
Vocab.	0.24†									
Speed	0.05	0.11								
Count	0.27*	0.48***	0.39**							
Flanker	0.32**	0.3*	0.42***	0.29*						
Stroop	0.3*	0.13	-0.01	0.16	0.25†					
Bear	0.11	0.24†	0.1	0.18	0.24†	0.24†				
DCCS	0.11	0.48***	0.32*	0.34**	0.45***	0.13	0.23†			
CEA	-0.06	0.09	0.08	0.16	-0.09	0.11	-0.03	0.29*		
CI	0.31*	0.53***	0.36**	0.4***	0.51***	0.33**	0.43***	0.32**	-0.1	
CFR	0.28*	0.41***	0.35**	0.5***	0.28*	0.11	0.26*	0.2	0.05	0.42***

Note. Simple correlation coefficients are shown; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; Speed = processing speed; †*p*<0.1; **p*<0.05; ***p*<0.01; ****p*<0.001.

MULTIPLE REGRESSION ANALYSES

Three-Year-Olds

A linear regression was used to determine whether EF accounted for a significant proportion of variance in children's causal reasoning in 3-year-olds. Child age, vocabulary, and processing speed were entered as covariates. The model accounted for a significant proportion of variance, $R^2 = 0.30$, $F(4, 647.17)$[1] = 10.45, $p < .001$. All variables except processing speed were significant predictors of causal reasoning (see Table 7).

[1] The F-test was calculated using the function micombine.F from the 'miceadds' package in R. The function uses a combination of F statistics for multiply imputed datasets using a chi square approximation.

Table 7. Regression Results for Three-Year-Olds Predicting the Causal Reasoning Composite Variable.

	Dependent variable: CR composite variable			
Variable	*B*	SE(*B*)	*t*-value	*p*-value
Age	0.25	(0.09)	2.70	0.01
Vocabulary	0.25	(0.09)	2.79	0.01
Processing Speed	0.07	(0.10)	0.68	0.50
EF	0.23	(0.09)	2.56	0.01

Note. $R^2 = 0.30$; *B* = unstandardized coefficients for standardized variables; standard errors are reported in parentheses; CR = causal reasoning; EF = executive function.

A second regression analysis was run with the cause-effect association task as the dependent variable. Child age, vocabulary, and processing speed were again entered as covariates. The model accounted for a small, but significant proportion of variance in the cause-effect association task, $R^2 = 0.12$, $F(4, 319.96)$[1] = 3.278, $p = 0.01$. Only processing speed was a significant predictor of cause-effect association (See Table 8).

Table 8. Regression Results for Three-Year-Olds Predicting the Cause-Effect Association Task.

	Dependent variable: Cause-effect association task			
Variable	*B*	SE(*B*)	*t*-value	*p*-value
Age	0.15	(0.10)	1.46	0.15
Vocabulary	-0.02	(0.10)	-0.19	0.85
Processing Speed	-0.35	(0.10)	-3.41	0.00
EF	0.12	(0.10)	1.17	0.24

Note. $R^2 = 0.12$; *B* = unstandardized coefficients for standardized variables; standard errors are reported in parentheses; EF = executive function.

Five-Year-Olds

A second set of linear regressions were used to determine whether EF accounted for a significant proportion of variance in 5-year-olds' causal reasoning skills. In the first model, age, vocabulary, and processing speed were entered as covariates, and our composite causal reasoning variable served as the dependent variable. The model accounted for significant proportion of variance, $R^2 = 0.54$, $F(4, 21.61)$[1] = 12.01, $p < .001$. All variables except child age were significant predictors (see Table 9).

Table 9. Regression Results for Five-Year-Olds Predicting the Causal Reasoning Composite Variable.

	Dependent variable: CR Composite Variable			
Variable	*B*	SE(*B*)	*t*-value	*p*-value
Age	0.15	(0.10)	1.50	0.14
Vocabulary	0.32	(0.11)	2.90	0.01
Processing Speed	0.26	(0.10)	2.56	0.01
EF	0.31	(0.12)	2.53	0.01

Note. $R^2 = 0.54$; *B* = unstandardized coefficients for standardized variables; standard errors are reported in parentheses; CR = causal reasoning; EF = executive function.

A second regression analysis was run with the cause-effect association task as the dependent variable. Child age, vocabulary, and processing speed were again entered as covariates. The model did not account for a significant proportion of variance in the cause-effect association task, $R^2 = 0.03$, $F(4, 88.24)$[1] = 0.34, $p = 0.85$. None of the variables emerged as significant predictors (see Table 10).

Table 10. Regression Results for Five-Year-Olds Predicting the Cause-Effect Association Task.

	Dependent variable: Cause-effect association task			
Variable	*B*	SE(*B*)	*t*-value	*p*-value
Age	-0.11	(0.16)	-0.69	0.50
Vocabulary	0.06	(0.16)	0.35	0.72
Processing Speed	-0.01	(0.16)	-0.08	0.94
EF	0.13	(0.17)	0.74	0.46

Note. R^2 = .03; *B* = unstandardized coefficients for standardized variables; standard errors are reported in parentheses; EF = executive function.

Chapter 4: Discussion

In this study, we examined whether EF has the potential to support the emergence of causal reasoning by evaluating patterns of correlation across a marked developmental transition. As hypothesized, we found that EF was a significant predictor of both 3- and 5-year olds' causal reasoning ability, as measured by a causal inference/counterfactual reasoning composite. Our results are in line with previous work that has shown that working memory capacity (Drayton, Turley-Ames, & Guajardo, 2011; Guajardo et al., 2009) and inhibition (Beck, Riggs, & Gorniak, 2009) in young children are related to their performance on counterfactual reasoning tasks. Similar relationships between EF and general reasoning abilities have also been shown in older children, adolescents (De Neys & Everaerts, 2008; Handley et al., 2004; Simoneau & Markovits, 2003), and young adults (e.g., Markovits & Doyon, 2004). However, these previous studies are limited in that they each only illuminate the relationship between one or two measures of EF (i.e., working memory or inhibition) and a single reasoning measure. This study is the first to our knowledge to systematically examine relations among three factors of EF (i.e., working memory, inhibition, and cognitive flexibility) and three distinct measures of causal reasoning (i.e., cause-effect association, causal inference and counterfactual reasoning) in a cross-sectional sample of young children.

In addition to confirming and elaborating upon earlier empirical work, our results are consistent with theoretical conceptions of EF and its relation to causal reasoning. For instance, as suggested by Gropen et al. (2011), the development of working memory might support the ability to compare representations of prior knowledge with observed evidence while also holding broad causal rules and potential actions in mind. Inhibition might also be important for helping children to focus on relevant information and revise previous beliefs in light of newly observed, and possibly conflicting, evidence (Gropen et al., 2011). In this way, children with stronger working memory and inhibition might more efficiently acquire causal knowledge and skills. Cognitive flexibility might also contribute by supporting children's ability to consider multiple possible courses of action and potential outcomes, as well as to apply and extend learned causal principles to novel circumstances. Recent work from Gopnik and colleagues (2017) specifically demonstrates how cognitive flexibility can enhance causal learning by keeping the child's mind open to new explanations, and therefore, novel insight.

Despite this apparent alignment of our findings with existing data and theory, the amount of variance accounted for by EF in the current study was actually substantially less than anticipated. In this context, it is important to note that although prior research has generally shown aspects of EF to be related to performance on causal reasoning tasks, a careful examination of the literature reveals some departures from this pattern. For example, Beck (2009) did not find a relationship between working memory and counterfactual reasoning (although they did detect a relationship between inhibition and counterfactual reasoning) in 3-4-year-old children. Likewise, Buchsbaum, Bridgers, Weisberg & Gopnik (2012), found no correlation between a stroop inhibition task and counterfactual reasoning in 3-4-year old children.

What might explain these patterns of results? One possibility is that causal reasoning is related to EF in more nuanced ways than previously thought. Although the bulk of the evidence suggests that EF is not yet differentiated into distinct componential skills in preschoolers, some have argued that a two-factor model best accounts for their performance (M. R. Miller et al., 2012). If true, some aspects of EF might be more influential than others, even at this early point in development. Indeed, some differential effects of this sort have been reported in the literature. For example, although Drayton et al. (2001) found a consistent relationship between inhibition and counterfactual reasoning across both 3- and 5-year-old age groups, working memory only played a significant role in explaining performance in the older age group. In the current study, we instead observed an increasing correlational relationship between inhibition and counterfactual reasoning with age. We might not have replicated Drayton et al.'s (2011) findings because we used a different measure of working memory which had a limited range, thereby potentially limiting its sensitivity to individual variability. Regardless, both findings hint at subtle complexities in the relationship between EF and causal reasoning that will require further attention in future research.

The relationship between early EF and causal reasoning might be further complicated by opposing influences, with some elements of EF supporting, and others interfering with, different aspects of causal reasoning performance. For example, while inhibition and cognitive flexibility together might be important for focusing a child's attention and suppressing prepotent responses in favor of considering alternative solutions and predictions, it is also possible that less inhibition might free children to fully exercise the flexibility to think creatively and consider additional possibilities from a larger hypothesis space (Gopnik et al., 2017). For example, in the tasks used in the current study, less inhibition might have allowed children to generate more creative or additional counterfactual solutions. Although intriguing, bringing full clarity to these potentially complex relationships is unfortunately beyond the scope of the current study. In future work, it will be important to add measures of divergent thinking to the slate of tasks under consideration to more directly test children's creatively.

Regardless of these nuances, the relatively weak relationship between EF and causal reasoning observed in the current work highlights the importance of keeping other factors in mind when considering the potential origins of causal reasoning. For example, vocabulary predicted significant variance in both age groups. This was not surprising given the heavy language demands of some of our tasks, and counterfactual reasoning in particular. A wealth of work has already established a strong relationship between counterfactual reasoning and language (e.g., Beck et al., 2009; Guajardo et al., 2009). The causal inference task also clearly relies on conceptual knowledge of objects, which one could argue goes hand-in-hand with vocabulary. Verbal ability might also be contributing in a deeper way to causal reasoning by providing a platform for new ways of thinking. For example, by scaffolding the transition from reflexive (i.e., reacting with impulsive responses) to reflective thought and action (i.e., pausing to consider multiple choices before acting) (Carlson & Moses, 2001). Verbal ability might also help children develop the ability to think abstractly, talk through larger problem spaces, and generate counterfactual scenarios.

Processing speed was also a significant predictor of causal reasoning at 5, but not 3, years of age. Processing speed is considered important for many aspects of thought and learning as it supports the ability to focus attention on, quickly discriminate between, and sequentially order, information. Indeed, processing speed was recently found to be significantly related to preschoolers' performance on EF tasks (Willoughby, Blair, Kuhn, & Magnus, 2018). Processing speed might be expected to have a particularly strong relation to tasks that include a reaction time component (e.g., flanker, DCCS). This pattern was in fact observed for the 5-year-olds, but not the 3-year-olds, in the current study. This is likely because the NIH-Toolbox incorporates reaction time into their scoring algorithm only after a participant achieves greater than 80% accuracy on more advanced levels of the tasks. Our 3-year-old participants were much less likely to reach this level of performance in the flanker and DCCS tasks, thereby leaving less opportunity for processing speed to influence their score. Note, however, that the influence of processing speed on causal reasoning might well transcend the parameters of any particular task. Faster processing speeds might help children more efficiently observe new data, retrieve relevant knowledge, and integrate these into potential solutions before key information is lost to interference or decay (Fry & Hale, 1996; Jensen, 1993; Kail & Ferrer, 2007; Kail & Park, 1994; L. T. Miller & Vernon, 1997). Consistent with this possibility, we observed strong correlations between processing speed and working memory in the current study at both 3 and 5 years of age.

But even when considered in total, the measures included in our models only accounted for 30 and 54% of the variance in causal reasoning (for 3- and 5-year-olds, respectively). This degree of unaccounted for variance could be taken as evidence that causal reasoning is a unique capability that is not wholly emergent from more fundamental cognitive skills, and therefore develops with some degree of independence. Before reaching this conclusion, however, it will be important to consider other potential contributors to the development of causal reasoning that were not assessed in this study. What might these be? One possibility is that causal reasoning is influenced by a child's causal stance, or their preference for, and interest in, causal information. In light of recent work demonstrating individual differences in preschoolers' causal stance (Alvarez & Booth, 2016), it is possible that children who are more drawn to and interested in causal information also become more practiced and proficient in reasoning causally, or are more motivated to engage in tasks that involve causal reasoning.

In order to fully understand the development of causal reasoning, it will also be important to consider environmental factors that might shape individual skills. For example, the degree to which a child engages in pretend play might be related to causal reasoning ability. Indeed, some have argued that pretend play, in and of itself, is a special case of counterfactual reasoning, as they both require considering alternatives that are inconsistent with reality (Amsel & Smalley, 2000; Dias & Harris, 1990; Guajardo & Turley-Ames, 2004; Kevin J. Riggs, Peterson, Robinson, & Mitchell, 1998; Sobel, 2006). In any case, pretend play is likely an important setting for practicing this skill in real-world contexts. The ability to engage in pretend play emerges during the second year of life, and is therefore certainly in place early enough to be a viable developmental precursor to causal reasoning (Fein, 1981; Leslie, 1987). In their exploration of this potential relationship, Buchsbaum et al. (2012) found that preschoolers made similar inferences about a causal system in both counterfactual and pretend play scenarios. Interestingly, they also found that pretend play accounted for significant variance in counterfactual reasoning above and beyond what was accounted for by EF. This experiential factor (i.e., practice with pretend play) might therefore be able to account for additional variance in causal reasoning that was not captured by EF and our covariates in the current work.

CAUSE-EFFECT ASSOCIATION TASK

Notably, despite replicating previously reported levels of performance on our measure of cause-and-effect association (e.g., Gopnik et al., 2001), this task was not associated with our other measures of causal reasoning (i.e., causal inference or counterfactual reasoning) in any of our analyses. This is surprising given that "blicket detector" tasks are often used as a classic demonstration of children's early causal reasoning abilities (e.g., Gopnik & Sobel, 2000; Kloos & Sloutsky, 2005; Sobel & Legare, 2014). One reason for the dissociation observed here might be that children performed at a high level on the cause-effect association task at both ages. This might indicate the task was too easy for the children in our sample, and therefore not sufficiently demanding on EF resources in a detectible way. Another reason might be that the cause-effect association task only required children to isolate arbitrary relationships between observable physical objects and events, while our causal inference and counterfactual reasoning tasks required that children focus on specific mechanisms to reason about meaningful causal relations. Moreover, while our cause-effect association task did require children to generate a novel intervention, it might not have tapped the complex Bayesian reasoning processes needed in similar tasks described in the literature (e.g., backward blocking or forward screening off). It is possible that a different version, which requires these more sophisticated Bayesian inferencing skills, might have been more tightly related to our other measures of causal reasoning. Consistent with this possibility, Weisberg and Gopnik (2013) specifically argue that the ability to think counterfactually (i.e., consider multiple possibilities, manipulate potential causes, and generate conclusions) is necessary to allow children to create and learn from causal Bayesian models (Gopnik & Tenenbaum, 2007; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Schulz, Kushnir, & Gopnik, 2007). However, because some evidence suggests that the ability to use counterfactual reasoning to scaffold learning new causal mechanisms does not emerge until the age of four (Schulz, Gopnik, & Glymour, 2007), this relationship might not be evident in the youngest children studied here.

Also contrary to our hypotheses, we found that EF was not a significant predictor of either 3- or 5-year olds' performance on the cause-effect association task. Again, these results might be because we lacked sufficient variance due to high levels of performance on the cause-effect association task. That said, some converging evidence that "blicket"-like tasks might not be integrally dependent on EF comes from a recent study that assessed groups of high and low socioeconomic status (SES) children on a more challenging "blicket detector" based causal learning task (Wente et al., 2017). Specifically, Wente and colleagues (2017) found that the low SES group performed similarly to their high SES peers on the causal learning task, despite lower EF scores. Although further research using a variety of "blicket detector" type tasks will be necessary to clarify these relationships, the current results suggest that the ability to detect cause-effect relations on the basis of arbitrary spatiotemporal contingencies is conceptually distinct from other aspects of causal reasoning, and does not rely heavily on EF skills.

LIMITATIONS

Although this study advances the field by exploring the relationship between EF and causal reasoning in preschoolers, there are clearly some limitations. For example, high rates of attrition and off-task behavior resulted in a large amount of missing data. We addressed this issue to the best of our ability by imputing complete and unbiased datasets for our regression analyses, but replication will be important for confirming the resulting analyses. From a methodological perspective, we were also limited in the number of causal reasoning measures both 3- and 5-year olds could successfully complete while skirting floor and ceiling effects, respectively. As a result, some of our measures (e.g., counterfactual reasoning and count and label) produced a restricted range of scores. Other tasks were too difficult for children (e.g., backward digit span) and contributed to the aforementioned missing data. Our small sample sizes also limited the analytic approaches available to us, specifically making it difficult to confidently interpret our structural equation model analysis (see Appendix).

Future studies with older children and additional measures will be crucial to further specify the relationship between more complex, and potentially more sensitive, measures of causal reasoning and EF. With this in mind, we are currently following the 3-year-olds tested here longitudinally through first grade. This will also allow us to address another limitation of this study, which is that we used a correlational design to examine relationships between EF and causal reasoning skills. A longitudinal design will allow us to explore possible moderation and mediation effects and also hold the effects of experiential factors (i.e., causal stance, pretend play, divergent thinking) constant. To the extent that EF and causal reasoning are malleable skills, intervention studies targeting these skills will serve to definitively test whether there is a causal relation.

CONCLUSIONS AND PRACTICAL IMPLICATIONS

In summary, this study found that EF significantly predicted causal reasoning performance across a period of rapid developmental change in both sets of skills from 3- to 5-years of age. However, a significant amount of variance in children's causal reasoning performance remained unaccounted for by EF and our other cognitive covariates (i.e., vocabulary and processing speed), thereby leaving open the possibility that some aspect of causal reasoning develops as a distinct capability. This was particularly true of the cause-effect association task, for which no association to EF (or to our other causal reasoning tasks) was detected. These results provide an important foundation for understanding the development of causal reasoning in the context of broader cognitive skills. As noted in the introduction, causal reasoning (i.e., making predictions, revising hypotheses) is likely a critical component of scientific literacy (Bauer & Booth, under review). Despite research demonstrating that achievement gaps in science begin to form before children even enter school (Greenfield et al., 2009), preschool classrooms typically do not focus on science. In part, achievement gaps in science may persist because we still know little about what knowledge and skills are fundamental to scientific literacy, and therefore would be most usefully targeted in early education. This study has begun to address this limitation by examining the developmental relationship between causal reasoning and EF, two possible underlying contributors to scientific understanding. As we build upon this foundational work, we will strive to clarify further the relationships between EF and causal reasoning skills, as well as the potential contributions of other factors (e.g., causal stance, pretend play). Indeed, we have begun to do so in the context of a longitudinal design, and hope ultimately to incorporate intervention methods to clarify whether observed relationships are causal in nature.

Appendix

STRUCTURAL EQUATION MODELING ANALYSES

We initially proposed to investigate the relationship between causal reasoning and EF using a structural equation modeling (SEM) framework. However, due to an unanticipated smaller sample size (especially for the 5-year-olds), we were unable to achieve adequate model fit, and thus could not calculate interpretable statistics using this approach. We therefore reported results from a simpler regression approach in the main body of this paper. For the sake of completeness, however, results from the originally planned SEM analyses are reported below.

Missing Data

First, to handle missing data in our multi-group SEM, we used the "lavaan" package that employs full information maximum likelihood estimation ("FIML"), allowing us to retain as many subjects as possible. "FIML" has been shown to produce unbiased parameter estimates and standard errors (Collins, Schafer, & Kam, 2001; Enders, 2006; Enders & Bandalos, 2001). In addition, we used maximum likelihood estimation with robust Huber-White standard errors ("MLR") and a scaled test statistic.

Assessing Model Fit

The χ^2 statistic provides a global index of model fit, where a non-significant χ^2 value suggests adequate model fit (Brown, 2014; Schumaker & Deshler, 2003).

Additional indexes used for model fit and comparison included the Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993), the Comparative Fit Index (CFI), and the Akaike Information Criterion (AIC). An RMSEA value less than 0.06 and CFI values above .95 indicate adequate fit (Hu & Bentler, 1999).

Multi-Group Structural Equation Model

In order to test whether the relationship between EF and causal reasoning measures shifted from 3- to 5-years of age, we first compared a multi-group structural equation model in which we constrained all the regression coefficients constrained to be equal to another model where the coefficients were allowed to vary freely. If the model where the coefficients were constrained to be equal fit the data better, then we could conclude there were no significant differences between the values of the regression coefficients between age groups. This would provide evidence that the development of EF (and our covariates) and CR might be parallel.

The model with all the regression parameters freely estimated fit the data poorly: CFI = .65; TLI = 0.53; RMSEA = 0.13, CI = 0.10, 0.15; AIC = 4567.54; SRMR = .14, according to fit criteria suggested by Hu and Bentler (1999). In addition, the overall chi-square was significant, $\chi^2(64) = 183.93$; $p < .001$. The fit indices for the constrained model also indicated a poor fit; CFI = .56; TLI = 0.42; RMSEA = 0.14, CI = 0.12, 0.17; AIC = 5013.05; SRMR= 0.15. In addition, the overall chi-square was significant, $\chi^2(64) = 211.12$; $p < .001$. Since the freely estimated model had a relatively better fit, this model was used for interpretation. The model showed that there were no significant differences in any of the regression coefficients between the 3- and 5-year-old age groups (see Table A1), though processing speed trended towards significance. Caution is required in interpreting these results due to the poor model fit observed.

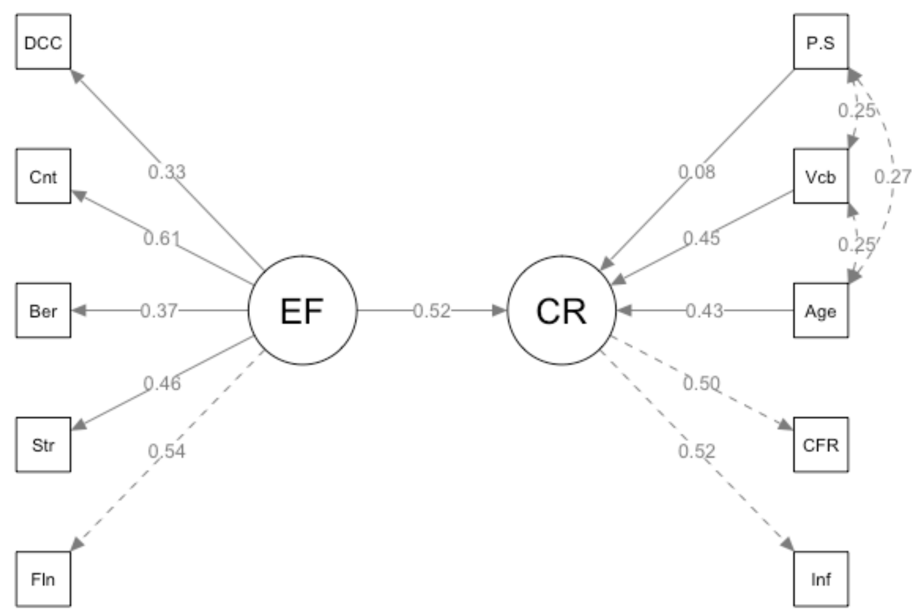
Table 11. Multiple-Group SEM Analysis.

	Dependent variable: CR latent variable				
Variable	Δ*B*	SE(*B*)	Δβ	*z*-value	*p*-value
Age	0.08	0.11	0.18	0.71	0.48
Vocabulary	-0.08	0.11	-0.09	-0.73	0.47
Processing Speed	-0.18	0.10	-0.31	-1.84	0.07
EF	0.02	0.47	-0.04	0.05	0.96

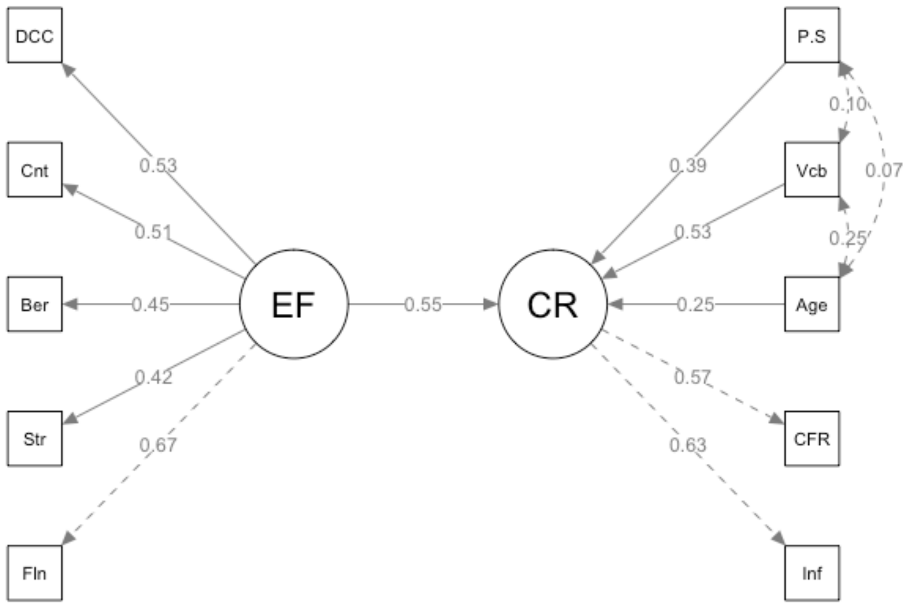
Note: This table shows the differences in regression coefficients between the 3- and 5-year-old age groups, with 3-year-olds serving as the reference group; CR = causal reasoning; EF = executive function.

Figure 4. Multiple Group SEM Diagrams for Freely Estimated Models.

A.



B.



Note. Panel A. Three-year-olds; Panel B. Five-year-olds; standardized coefficients are shown; CR = causal reasoning; EF = executive function; Fln = Flanker; Str = stroop; DCC = DCCS; Ber = bear/dragon; Cnt = count/label; Inf = causal inference, CFR = counterfactual reasoning; P.S. = processing speed; Vcb = vocabulary.

Analysis of Supplementary Cognitive Measures

We initially proposed to examine additional measures of working memory including measures from the Wechsler Preschool and Primary Scale of Intelligence Fourth Edition and forward and backward digit span. However, after further reflection and consultation with the literature, we determined that these measures (with the exception of backward digit span), actually tapped short-term memory rather than working memory. Because we did not have any hypotheses regarding the relationship between children's short-term memory and causal reasoning ability, we did not include these measures in our primary analyses. Even though the backward digit span task did tap working memory, we chose to exclude it from our primary analyses as well because the majority of children (especially from our 3-year-old sample) could not successfully complete even the practice trials. Nevertheless, descriptions, and correlations between, these measures and our primary measures are reported below.

Measures

Wechsler Preschool and Primary Scale of Intelligence (Fourth Edition)

Two subtests from the Wechsler Preschool 4th Edition (WPPSI-IV; Wechsler, 2012) were used to assess short-term memory. In the Picture memory subtest, participants viewed one or more pictures for a specified time and then select these pictures within a field of distracting pictures on the response page. The Picture Memory subtest measures visual working memory using the familiarize–recognize paradigm, for which a set of stimuli is viewed and then recognized from among a set of responses. In the Zoo Location subtest, participants viewed one or more animal cards placed on zoo map for a limited time and then must place each card in the previously viewed location. The Zoo Locations subtest measures visual spatial working memory using the observe–perform paradigm, wherein some action or actions are observed and then repeated. The dependent variable in Zoo Locations is an age adjusted and scaled score based on the raw number of trials competed correctly.

Forward and Backward Digit Span

In this task (Davis & Pratt, 1995), the experimenter introduced children to a puppet named Mr. Bear, a friendly puppet who wants to play a number game. Children first completed forward digit span as a warm up. For forward digit span, children were instructed to repeat the numbers Mr. Bear said. Two practice trials were administered and feedback was given if children misunderstood the rules. The experimenter ended the task and moved on to backwards digit span when children failed two consecutive items within a trial block. For backwards digit span, the experimenter began the task by saying, "Mr. Bear has another number game to play! Now, instead of repeating after Mr. Bear, now you're going to say what he says backwards!" The experimenter demonstrated by making Mr. Bear say "1, 2." She then modeled the correct answer by saying, "2, 1." Children were then invited to try (using the same example) and another practice trial. The practice and next four trials had two digits, and then the number of digits increased until children answered incorrectly on two consecutive items within a trial block.

Descriptive Statistics

The means, standard deviations, and ranges for measures of short-term memory are shown in Tables A2 and A4. Simple bivariate correlations between these variables of interest are presented in Tables A3 and A5. Please note that we do not report scores for backward digit span for the 3-year-old group as the majority of children were unable to successfully pass the practice trials on this task.

For the 3-year-olds, all three measures of short-term memory (WPPSI-picture memory, WPPSI-zoo locations and digit span) correlated with counterfactual reasoning performance. Digit span also correlated significantly with performance on the causal inference tasks. No correlations to the cause-effect association task were observed.

Table 12. Descriptive Statistics for Supplemental Short-Term Memory Measures for Three-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
WPPSI-IV PM	10.52	4.89	0	21
WPPSI-IV ZL	7.38	2.68	0	12
Forward Digit Span	3.50	1.76	0	7

Note. WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations.

Table 13. Correlations Between Short-Term Memory Variables for Three-Year-Old Data After Imputation.

	WPPSI PM	WPPSI ZL	Digit Span
Age	0.3**	0.14	0.22*
Gender	0.13	-0.02	-0.01
Vocab.	0.34***	0.21*	0.3**
Speed	0.35**	0.23*	0.26**
Count	0.23*	0.38**	0.34***
Flanker	0.36***	0.34***	0.32***
Stroop	0.01	0.25*	0.17
Bear	0.15	0.21*	$0.21^{\dagger}$
DCCS	0.27*	0.15	0.14
CEA	0.07	0.04	-0.05
CI	0.11	$0.19^{\dagger}$	0.32**
CFR	0.32**	0.23*	0.25*
WPPSI PM		0.4***	0.25*
WPPSI ZL			$0.19^{\dagger}$

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; $^{\dagger}p<0.1$; $*p<0.05$; $**p<0.01$; $***p<0.001$.

For the 5-year-olds, no correlations between short-term memory and causal reasoning measures were evident. However, the one supplemental measure of working memory included here (i.e., backward digit span), correlated significantly with performance on the causal inference task.

Table 14. Descriptive Statistics for Supplemental Short-Term Memory Measures for Five-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
WPPSI-IV PM	15.75	5.42	2	28
WPPSI-IV ZL	9.68	2.56	1	13
Forward Digit Span	5.75	1.55	2	9
Backward Digit Span	3.60	2.07	0	7

Note. WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations.

Table 15. Correlations Between Short-Term Memory Variables for Five-Year-Old Data After Imputation.

	WPPSI PM	WPPSI ZL	Digit Span	Backward Digit Span
Age	0.07	0.04	0.17	0.3*
Gender	-0.05	0.1	-0.03	-0.04
Vocab.	0.42***	0.01	0.27*	0.34**
Speed	0.26*	0.16	0.04	0.34**
Count	0.29*	-0.1	0.25†	0.38**
Flanker	0.18	0.17	0.24†	0.41**
Stroop	0.05	0.02	0.26†	0.36**
Bear	0.18	0.07	0.02	-0.01
DCCS	0.35**	-0.01	0.27*	0.25*
CEA	-0.04	-0.17	0.24†	0.17
CI	0.18	0.12	0.1	0.25*
CFR	0.24†	-0.17	0.1	0.23†
WPPSI PM	-	0.31*	0.16	0.19
WPPSI ZL		-	-0.09	-0.09
Digit Span			-	0.44***

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; †p<0.1; *p<0.05; **p<0.01; ***p<0.001.

ANALYSIS OF SUPPLEMENTAL PARENT REPORT MEASURES

Although in our original research plan, we intended to consider parent report measures of child characteristics along with our direct assessments, after consultation with experts and further review of the literature, we determined these measures were not appropriate for our purposes. Specifically, we learned that they were intended for clinical purposes and were unlikely to be sufficiently sensitive to detect subtle variations within a sample of typically developing children. The fact that we removed children from our analyses who exhibited poor behavior and attention during our assessments might well have further stripped the sensitivity of these measures, thereby diminishing their usefulness for our purposes. Nevertheless, we present descriptions of these measures, and their correlations to our key factors of interest below.

Measures

Behavior Rating Inventory of Executive Function Preschool Version

Parents independently completed the Behavior Rating Inventory of Executive Function Preschool Version (BRIEF-P; Gioia, Isquith, Retzlaff, & Espy, 2002), a parent report of executive function. This survey evaluates five aspects of executive functioning including Inhibition, Shifting, Emotional Control, Working Memory, and Planning/Organizing. Total time to complete this survey was approximately 10 minutes. The outcome variable obtained is a standardized global executive percentile score.

Children's Behavior Questionnaire

The Children's Behavior Questionnaire (CBQ; Putnam, 2006) is an assessment of temperament intended for use with preschool aged children. The CBQ Very Short form consists of 36 items. The CBQ was administered by a trained research assistant or completed independently by a parent. Parents rated children's behaviors on a likert scale ranging from 1-7 (where 1 corresponds to "extremely untrue of your child," and "extremely untrue of your child". Total administration time was approximately 10 minutes. This task was scored by taking the means of items relevant to each sub-domain including Negative Affectivity, Surgency, and Effortful Control.

Descriptive Statistics

The means, standard deviations, and ranges for supplemental parent survey data are shown in Tables A6 and A7. Simple bivariate correlations between these variables of interest are presented in Tables A8 and A9.

Table 16. Descriptive Statistics for Parent Report Measures for Three-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
CBQ Surgency	4.45	0.79	2.17	6.58
CBQ Effortful Control	5.23	0.70	2.75	6.42
BRIEF-P Global Score	46.93	27	0	99

Note. BRIEF-P Global Executive Score is a percentage.

Table 17. Descriptive Statistics for Parent Report Measures for Five-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
CBQ Surgency	4.46	0.84	2.58	6.25
CBQ Effortful Control	5.57	0.64	4.25	6.83
BRIEF-P Global Score	38.69	28.43	0	99

Note. BRIEF-P Global Executive Score is a percentage.

Table 18. Correlations Between Parent Report Measures for Three-Year-Old Data After Imputation.

	CBQ Surg.	CBQ Eff.	BRIEF
Age	-0.13	0.06	0
Gender	0.08	-0.11	-0.05
Vocab.	0.02	0.27**	0.23*
Speed	0.14	-0.03	-0.05
Count	-0.1	0.02	0
Flanker	-0.14	0.17†	0.18†
Stroop	-0.1	-0.07	0.03
Bear	0.09	0.1	0.04
DCCS	-0.33***	0.25*	-0.1
CEA	-0.11	0.12	-0.01
CI	0	0.11	0.23*
CFR	0.07	0.15	0.06
WPPSI PM	-0.07	0.13	0.17
WPPSI ZL	-0.2*	0.06	0.22*
Digit Span	0.02	0.12	0.36***
CBQ Surg.		-0.18†	0.11
CBQ Eff.			0.04

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; †p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table 19. Correlations Between Parent Report Variables for Five-Year-Old Data After Imputation.

	CBQ Surg.	CBQ Eff.	BRIEF
Age	0.22[†]	0.06	0.09
Gender	0.06	-0.4***	0.02
Vocab.	0.05	-0.1	0.18
Speed	0.12	0.16	-0.12
Count	0.06	0	-0.03
Flanker	0.17	0.3*	-0.07
Stroop	0.03	0.26*	0.03
Bear	0.03	0.21	-0.04
DCCS	0.06	0.26*	0.01
CEA	0.11	0.01	-0.06
CI	0.15	0.18	0.17
CFR	0.17	0.12	-0.1
WPPSI PM	-0.02	0.04	0.03
WPPSI ZL	-0.02	-0.07	0.02
Digit	-0.08	0.21	0.15
Backward Digit	0	0.02	-0.04
CBQ Surg.	-	0.06	-0.04
CBQ Eff.		-	-0.15

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; [†]$p<0.1$; *$p<0.05$; **$p<0.01$; ***$p<0.001$.

ANALYSIS OF SUPPLEMENTAL ENVIRONMENTAL FACTORS

Here we also present correlations between an additional parent survey of home environment, as well as maternal education, and our primary variables. Although these measures were not a core component of our proposed plan, they may be enlightening with respect to individual differences in causal reasoning ability. Relevant descriptions of measures and correlations with our other key measures are reported below.

Measures

StimQ-Preschool

The StimQ-Preschool (StimQ; Mendelsohn, Dreyer, Tamis-LeMonda, & Ahuja, 1999) measures the cognitive home environment across the preschool period. The form consists of four subscales: Availability of Learning Materials, Reading, Parental Involvement in Developmental Advance, and Parental Verbal Responsivity. The outcome variable obtained is a standardized total score.

Maternal Education

Parents reported their highest degree obtained. We categorized maternal education into 5 discrete categories: (1) no high school degree, (2) high school degree, (3) some college or additional training beyond high school, (4) four-year bachelor's degree, (5) master's degree or higher. Very few parents (2.71%) declined to report the mother's highest level of education.

Descriptive Statistics

The means, standard deviations, and ranges for the StimQ and maternal education are shown in Tables A10 and A11. Simple bivariate correlations between these variables of interest are presented in Tables A12 and A13. Of note, the StimQ was significantly correlated with causal inference and marginally correlated with counterfactual reasoning for the 3-year-olds, but not for 5-year-olds. Maternal education was not correlated with any of the causal reasoning measures for either group.

Table 20. Descriptive Statistics for Environmental Measures for Three-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
StimQ Total Score	42.22	4.36	29	49
Maternal Education	3.16	1.45	0	5

Table 21. Descriptive Statistics for Environmental Measures for Five-Year-Old Data After Imputation.

Variables	Mean	SD	Min	Max
StimQ Total Score	44.26	4.76	20	49
Maternal Education	3.09	1.48	1	5

Table 22. Correlations Between Environmental Measures for Three-Year-Old Data After Imputation.

	StimQ	Mat. Ed.
Age	0.27**	0.02
Gender	0.12	-0.01
Vocab.	0.31**	0.15
Speed	0.19†	0.21*
Count	0.24*	0.24*
Flanker	0.13	-0.03
Stroop	0.05	0.05
Bear	0.16	0.12
DCCS	0.26**	-0.03
CEA	-0.02	-0.09
CI	0.38***	-0.05
CFR	0.17†	0.08
WPPSI PM	0.31**	0.15
WPPSI ZL	0.15	0.08
Digit Span	0.18†	0.28**
CBQ Surg.	-0.01	0.11
CBQ Eff.	0.22*	-0.06
BRIEF	0.01	-0.14
StimQ		0.12

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; $^{\dagger}p<0.1$; $*p<0.05$; $**p<0.01$; $***p<0.001$.

Table 23. Correlations Between Environmental Variables for Five-Year-Old Data After Imputation.

	StimQ	Mat.Ed.
Age	0.01	-0.08
Gender	-0.11	-0.22†
Vocab.	-0.1	0.03
Speed	-0.05	0.1
Count	0.15	0.25*
Flanker	-0.07	0.2
Stroop	0.22	0.27*
Bear	0.13	0.32**
DCCS	-0.11	0.09
CEA	-0.11	0.1
CI	0.02	0.14
CFR	-0.07	0.1
WPPSI PM	0.22	0.17
WPPSI ZL	0.05	-0.04
Digit	0.24†	0.27*
Backward Digit	0.2	0.14
CBQ Surg.	-0.17	-0.12
CBQ Eff.	0.18	0.23†
BRIEF	0.11	0.02
StimQ		0.31*

Note. Simple correlation coefficients are shown; count = count/label; bear = bear/dragon; speed = processing speed; CEA = cause-effect association; CI = causal inference; CFR = counterfactual reasoning; WPPSI PM = WPPSI-IV picture memory; WPPSI ZL = WPPSI-IV zoo locations; CBQ Surg. = CBQ Surgency; CBQ Eff. = CBQ Effortful Control; Mat. Ed. = maternal education; $^{\dagger}p<0.1$; $*p<0.05$; $**p<0.01$; $***p<0.001$.

References

Alvarez, A., & Booth, A. E. (2014). Motivated by Meaning: Testing the Effect of Knowledge-Infused Rewards on Preschoolers' Persistence. *Child Development, 85*(2), 783-791. doi:10.1111/cdev.12151

Alvarez, A., & Booth, A. E. (2016). Exploring individual differences in preschoolers' causal stance. *Developmental Psychology, 52*(3), 411-422. doi:10.1037/dev0000085

Amsel, E., & Smalley, J. D. (2000). Beyond really and truly: Children’s counterfactual thinking about pretend and possible worlds. *Children’s reasoning and the mind, 22*(5), 121-147.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology, 8*(2), 71-82.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89): Elsevier.

Baddeley, A. D., & Hitch, G. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485.

Bauer, J. R., & Booth, A. E. (under review). Exploring Potential Cognitive Foundations of Scientific Literacy in Preschoolers: Causal Reasoning and Executive Function. *Early Childhood Research Quarterly*.

Beck, S. R., Burns, P., & Riggs, K. J. (2011). Multiple Developments in Counterfactual Thinking. In *Understanding Counterfactuals, Understanding Causation*. Oxford: Oxford University Press.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning, 15*(4), 337-354. doi:10.1080/13546780903135904

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327-336.

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*(2), 647--663. doi:10.1111/j.1467-8624.2007.01019.x

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. E. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322-330.

Brand, J. (1999). *Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets*.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of "hot" and "cool" executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*(3), 337--349. doi:10.1016/j.ecresq.2009.06.001

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*: Guilford Publications.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions, 154*, 136-136.

Buchsbaum, D., Bridgers, S., Weisberg, D., & Gopnik, A. (2012). The power of possibility: causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2202-2212. doi:10.1098/rstb.2012.0122

Bullock, M., & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development, 50*(1), 89-96.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In *The developmental psychology of time* (pp. 209-254). New York: Academic Press.

Byrne, R. M. (2007). *The rational imagination: How people create alternatives to reality*: MIT press.

Carlozzi, N. E., Tulsky, D. S., Kail, R. V., & Beaumont, J. L. (2013). VI. NIH Toolbox Cognition Battery (CB): measuring processing speed. *Monographs of the Society for Research in Child Development, 78*(4), 88-102.

Carlson, S. M. (2005). Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology, 28*(2), 595 - 616.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods, 6*(4), 330.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition, 120*(3), 341-349.

Das Gupta, P., & Bryant, P. E. (1989). Young Children's Causal Inferences. *Child Development, 60*(5), 1138-1146. doi:10.2307/1130788

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*(1), 25-31. doi:10.1080/00049539508258765

De Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology, 100*(4), 252-263.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*(1), 637-676.

Diamond, A. (2002). Faculty of 1000 evaluation for Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. In.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*(1), 135--168. doi:10.1146/annurev-psych-113011-143750

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool Children's Performance in Task Switching on the Dimensional Change Card Sort Task: Separating the

Dimensions Aids the Ability to Switch. *Developmental Neuropsychology, 28*(2), 689 - 729.

Dias, M., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8*(4), 305-318.

Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology, 108*(3), 532--548. doi:10.1016/j.jecp.2010.09.007

Enders, C. K. (2006). Analyzing structural equation models with missing data. *Structural equation modeling: A second course*, 313-342.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457.

Espy, K. A., Kaufmann, P. M., McDiarmid, M. D., & Glisky, M. L. (1999). Executive Functioning in Preschool Children: Performance on A-Not-B and Other Delayed Response Format Tasks. *Brain and Cognition, 41*(2), 178-199.

Fein, G. G. (1981). Pretend play in childhood: An integrative review. *Child development*, 1095-1118.

Fernbach, P. M., Macris, D. M., & Sobel, D. M. (2012). Which one made it go? The emergence of diagnostic reasoning in preschoolers. *Cognitive Development, 27*(1), 39-53. doi:10.1016/j.cogdev.2011.10.002

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*(4), 237-241.

Frye, D., Zelazo, P. D., Brooks, P. J., & Samuels, M. C. (1996). Inference and Action in Early Causal Reasoning. *Developmental Psychology January, 32*(1), 120-131.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483-527. doi:10.1016/0885-2014(95)90024-1

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31--60. doi:10.1037/0033-2909.134.1.31

Gelman, R., Bullock, M., & Meck, E. (1980). Preschoolers' understanding of simple object transformations. *Child development*, 691-699.

Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., . . . Weintraub, S. (2013). IV. NIH Toolbox Cognition Battery (CB): Measuring Language (Vocabulary Comprehension and Reading Decoding). *Monographs of the Society for Research in Child Development, 78*(4), 49--69. doi:10.1111/mono.12034

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology, 8*(4), 249-257.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299-323).

Gopnik, A., & Glymour, C. (2002). Causal maps and Bayes nets: A cognitive and computational account of theory-formation. In *Carruthers, Peter (Ed); Stich,* *Stephen (Ed); et al (2002) The cognitive basis of science* (pp. 117-132). New York, NY: Cambridge University Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review, 111*(1), 3-32.

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., . . . Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences, 114*(30), 7892-7899. doi:10.1073/pnas.1700811114

Gopnik, A., & Schulz, L. E. (2007). *Causal learning: Psychology, philosophy, and computation*. New York, NY: Oxford University Press.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 1205-1222.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*(5), 620-629.

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science, 10*(3), 281-287.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the Preschool Classroom: A Programmatic Research Agenda to Improve Science Readiness. *Early Education and Development, 20*(2), 238-264. doi:10.1080/10409280802595441

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What Do Children Want to Know About Animals and Artifacts? Domain-Specific Requests for Information. *Psychological Science, 17*(6), 455-459.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences, 14*(8), 357-364.

Gropen, J., Clark-Chiarelli, N., Hoisington, C., & Ehrlich, S. B. (2011). The Importance of Executive Function in Early Science Education. *Child Development Perspectives, 5*(4), 298-304. doi:10.1111/j.1750-8606.2011.00201.x

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology, 27*(3), 681--702. doi:10.1348/026151008X357886

Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development, 19*(1), 53-80. doi:10.1016/j.cogdev.2003.09.002

Hagmayer, Y., Sloman, S., Lagnado, D., & Waldmann, M. R. (2007). Causal reasoning through intervention. In *Causal learning: Psychology, philosophy, and computation* (pp. 86-100). New York, NY: Oxford University Press.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). : Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195.

Hommel, B. (2011). Attention and spatial stimulus coding in the Simon task: a rejoinder to van der Lubbe and Abrahamse (2010). *Acta psychologica, 136*(2), 265-268.

Hong, L., Chijun, Z., Xuemei, G., Shan, G., & Chongde, L. (2005). The influence of complexity and reasoning direction on children's causal reasoning. *Cognitive Development, 20*(1), 87-101.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233-253.

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of experimental child psychology, 108*(3), 663-676.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017--2036. doi:10.1016/j.neuropsychologia.2006.01.010

Jensen, A. R. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17*(1), 47-77.

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development? *Developmental cognitive neuroscience, 1*(1), 7-21.

Kail, R., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child Development, 78*(6), 1760-1770.

Kail, R., & Park, Y.-S. (1994). Processing Time, Articulation Time, and Memory Span. *Journal of Experimental Child Psychology, 57*, 281-291.

Kemler Nelson, D. G., Chan Egan, L., & Holt, M. B. (2004). When Children Ask "What Is It?" What Do They Want to Know About Artifacts? *Psychological Science, 15*(6), 384-389.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449-467.

Kloos, H., & Sloutsky, V. M. (2005). *Do preschoolers understand causality? A critical look.* Paper presented at the Proceedings of the Cognitive Science Society.

Kun, A. (1978). Evidence for preschoolers' understanding of causal direction in extended causal sequences. *Child Development*, 218-222.

Kushnir, T., Gopnik, A., Schulz, L., & Danks, D. (2003). Inferring hidden causes. In *Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society.* Boston, MA.: Cognitive Science Society.

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental Changes in Executive Functioning. *Child Development, 84*(6), 1933-1953. doi:10.1111/cdev.12096

Legare, C. H. (2012). Exploring Explanation: Explaining Inconsistent Evidence Informs Exploratory, Hypothesis-Testing Behavior in Young Children. *Child Development, 83*(1), 173-185. doi:10.1111/j.1467-8624.2011.01691.x

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning. *Child Development, 81*(3), 929-944.

Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*(1), 59-80.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review, 94*(4), 412.

Marcis, D. M., & Sobel, D. M. (2017). The Role of Evidence Diversity and Explanation in 4- and 5-year-Olds' Resolution of Counterevidence. *Journal of Cognition and Development, 18*(6), 358-374. doi:http://dx.doi.org/10.1080/15248372.2017.1323755

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & cognition, 32*, 592-601.

McCormack, T., & Hoerl, C. (2005). Children's reasoning about the causal significance of the temporal order of events. *Developmental Psychology, 41*(1), 54.

McCormack, T., & Hoerl, C. (2007). Young children’s reasoning about the order of past events. *Journal of experimental child psychology, 98*(3), 168-183.

Mendelsohn, A. L., Dreyer, B. P., Tamis-LeMonda, C. S., & Ahuja, P. (1999). Validity of StimQ, a Scale for Assessing the Cognitive Home Environment. *Journal of Developmental & Behavioral Pediatrics, 20*(5), 399.

Miller, L. T., & Vernon, P. A. (1997). Developmental Changes in Speed of Information Processing in Young Children. *Developmental Psychology May, 33*(3), 549-554.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*(3), 395-423.

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*(1), 49-100.

Müller, U., Zelazo, P. D., Hood, S., Leone, T., & Rohrer, L. (2004). Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development, 75*(5), 1594-1609.

Putnam, S. a. R., Mary. (2006). Development of Short and Very Short Forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102-112. doi:10.1207/s15327752jpa8701_09

Rafetseder, E., & Perner, J. (2014). Counterfactual reasoning: sharpening conceptual distinctions in developmental studies. *Child development perspectives, 8*(1), 54-58.

Rafetseder, E., Renate, C.-V., & Perner, J. (2010). Counterfactual Reasoning: Developing a Sense of "Nearest Possible World". *Child Development, 81*(1), 376-389.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology, 114*(3), 389-404. doi:10.1016/j.jecp.2012.10.010

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly, 30*(2), 131-147.

Riggs, K. J., & Beck, S. R. (2007). Thinking developmentally about counterfactual possibilities. *Behavioral and Brain Sciences, 30*(5-6), 463-463.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*(1), 73-90.

Robinson, E. J., & Beck, S. (2000). What is difficult about counterfactual reasoning. *Children’s reasoning and the mind*, 101-119.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43*(5), 1124-1139.

Schulz, L. E., & Gopnik, A. (2004). Causal Learning Across Domains. *Developmental Psychology, 40*(2), 162-176.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*(3), 322--332. doi:10.1111/j.1467-7687.2007.00587.x

Schulz, L. E., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Intervention and causal inference. In *Causal learning: Psychology, philosophy, and computation* (pp. 67-85). New York, NY: Oxford University Press.

Schumaker, J. B., & Deshler, D. D. (2003). *Designs for applied educational research*: Swanson, H. Lee (Ed); Harris, Karen R (Ed); Graham, Steve (Ed).

Shultz, T. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, 1-51.

Shultz, T., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, 394-399.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*(6), 964.

Sobel, D. M. (2006). How fantasy benefits young children's understanding of pretense. In (Vol. 9, pp. 63-75).

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*(6), 1103.

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science, 10*(3), 286-306.

Sobel, D. M., & Legare, C. H. (2014). Causal learning in children. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(4), 413--427. doi:10.1002/wcs.1291

Sobel, D. M., & Munro, S. E. (2009). Domain Generality and Specificity in Children's Causal Inference About Ambiguous Data. *Developmental Psychology, 45*(2), 511-524.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303-333.

Sophian, C., & Huber, A. (1984). Early developments in children's causal judgments. *Child Development*, 512-526.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745--759. doi:10.1080/17470210500162854

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive sciences, 10*(7), 309-318.

Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of statistical computation and simulation, 76*(12), 1049-1064.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *2011, 45*(3), 67. doi:10.18637/jss.v045.i03

Wechsler, D. (2012). Wechsler Preschool and Primary Scale of Intelligence - fourth edition technical manual and interpretive manual. In. San Antonio, TX.: Psychological Corporation.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*(11 Supplement 3), S54-S64. doi:10.1212/WNL.0b013e3182872ded

Weisberg, D. S., & Gopnik, A. (2013). Pretense, counterfactuals, and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37*(7), 1368-1381.

Wente, A. O., Kimura, K., Walker, C. M., Banerjee, N., Fernández Flecha, M., MacDonald, B., . . . Gopnik, A. (2017). Causal learning across culture and socioeconomic status. *Child development*.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575--587. doi:10.1037/0012-1649.44.2.575

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*(3), 436-452.

Wilkening, F., & Sodian, B. (2005). Scientific Reasoning in Young Children: Introduction. *Swiss Journal of Psychology, 64*(3), 137-139. doi:10.1024/1421-0185.64.3.137

Willoughby, M. T., Blair, C. B., Kuhn, L. J., & Magnus, B. E. (2018). The benefits of adding a brief measure of simple reaction time to the assessment of executive function skills in early childhood. *Journal of experimental child psychology, 170*, 30-44.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development, 85*(1), 160-175.

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

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. *Language structure, discourse, and the access to consciousness, 12*, 113-153.