Representing Minds Representing Minds: An Examination of the Association between Recursive Mental State Attributions and Executive Processes

by

Claudie M. Peloquin

A Thesis Submitted to Victoria University of Wellington in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

In the

School of Psychology

Victoria University of Wellington

2021

Abstract

Theory of Mind (ToM) and Executive Functioning (EF) are two pillars of human social cognition often studied in conjunction, but rarely considered together beyond childhood. Adults routinely undertake ToM activities of higher levels, such as those that require reasoning recursively through other individuals' presumed reasoning about others (e.g., she believes that he believes that this is difficult to grasp). The possibility of links between EF and these special kinds of representations, termed second-order ToM, is explored for the first time in the work presented herein, and documented in two ways. First, the research plan and hypotheses at the basis of this dissertation were informed in large part by a meta-analytic review of the extant literature linking EF to second-order ToM (Study 1). We employed multilevel modelling techniques to estimate the pooled effect size of over 80 correlation coefficients, extracted from both school-age children and adult samples ($N = 2584$). While the developmental literature provided evidence of second-order ToM-EF linkage in children, the adult findings were weaker and more difficult to interpret for a variety of methodological reasons. Hence, in Studies 2 and 3, we introduced a new age-adapted methodological paradigm, and conducted an extensive mapping of the adult capacity for second-order ToM reasoning in relation to EF. Across the two studies, we found that individual differences in both working memory and cognitive flexibility correlated with second-order ToM performance, irrespective of variance accounted for by factors such as verbal ability (Study 2) and non-verbal ability (Study 3). In Study 3 we explored the relation between second-order ToM and EF with greater specificity. We replicated the findings of Study 2 and found that the manipulation component of working memory (but not the storage component) and the task-switching component of cognitive flexibility (but not the set-shifting component) were central to the relation. The broader theoretical implications of our findings

were then discussed, along with suggestions for potential ways forward in the study of the relation between second-order ToM and EF in adulthood.

Dedication

In memory of Katie Hunter (1991-2018), a kind-hearted, cheerful, and incredibly strong woman.

Acknowledgements

I am deeply indebted to a great many people who have supported me throughout this PhD journey. Foremost, I am extremely grateful to my academic advisor, Dr Lou Moses. This dissertation would never have been possible without Lou's incredible mentoring, and I truly hope that my future contributions to the field will be tainted by what I have learned working under Lou's supervision: kindness, patience and support are everything. I would also like to thank my co-advisor, Dr Alia Martin, for her invaluable guidance and scholarly inputs, as well as Dr Jason Low for his advice on earlier research work. I also extend my thanks to the support staff at Victoria University of Wellington and the faculty of graduate research with special mentions to Dr Sara Cottrell, Dr Matt Crawford and Patricia Stein for being so generous of their time and ensuring that I reached this point in good health and academic standing.

To my incredibly talented, intelligent fellow labmates, I could not have asked for a better support system than what I have been extremely lucky to find in each of you living in EA410. Katheryn, you have been a role model, a friend, a sister and an insatiable source of inspiration. Schyana, I am forever grateful for your genuine kindness, calming serenity, and reassuring presence. Chris, thank you for making yourself complicit in my obsession for what others think that I think that they think that you think. Catherine, your input has been invaluable in sharpening the ideas of this thesis (as were the numerous stairwell rants and rambles!). Cong, Pieter, Giovanni, Reneeta and Radhika, you all have contributed to my sanity by providing honest, frank and authentic advice at times when I most needed. Many thanks to the VuW BabyLab and UOregon DML North, especially to Laura, Mary Beth, Michaela, Karlena and Stephanie (and Reagan!) for their thorough and insightful feedback. Sam Mohd Akhir, Al Abenoja and Braden Cradock, thank you for all of the 1096548 times you came to my rescue with IT stuff, and for handling my tech "emergencies" with a level of patience and kindness I most certainly did not always deserve.

I am also particularly grateful to a few special people who deserve my eternal recognition for putting up with me through thick and thin. Raphaelle, Erika, Caroline, Amelie, Myrsini, Rebecca, Robinson, Nihit, Antonin, Valle, Marcus, Daniel, Misa, Marcel, Laureline, Wouter, Maja and Rahel, thank you for steering me back to reality at times when the scope and size of this project seemed overwhelming, ridiculous and impossible. You all have been remarkable proofreaders, counsellors, living room dance partners, cheerleaders and lactose-free-everything providers. Thank you for celebrating with me even the slightest of milestones, and for never letting me forget that there is so much more to life than academics. Dr David, words cannot express how truly grateful I am for the countless evening strolls, deep meaningful (!) kitchen chats, chocolate/wine therapies and overall support and friendship. Finally, many thanks go to my family and friends on the other side of the world, for their encouragement, support and understanding. None of this would have been possible without their unconditional love and faith that I can do anything I set my mind to. Merci de croire en moi depuis le jour 1!

Table of Contents

List of Tables

List of Figures

CHAPTER 1

General Introduction

Beliefs, and especially false-beliefs, are special entities. They occupy a central role in our understanding of the means-end process by which actions are generated and constrained, and help us explain observed behaviour in rational terms. For instance, I can reasonably explain why someone would reach for an empty cookie jar by reflecting upon her beliefs in that situation: She wants to eat cookies and believes that there are cookies in the cookie jar. I understand, by means of inferring her unseen mental states, that her actions were conditioned by what she believed to be true in this context, and made sense according to her own (mis)representation of reality. By doing so, I performed one of the most remarkable features of human cognition: I recognized this person as independently minded, with her own attitudes toward objects or aspects of a given situation (Dennett, 1987; Leslie, 1994; Premack & Woodruff, 1978).

If I use this ability recursively, to attribute not only a mind but a Theory of Mind (ToM) to another individual, the inferences that I will need to make in anticipation of their behaviour will need to account for the beliefs that they, themselves, have already ascribed to yet another individual. I will need to reason about ToM itself, in order to think through multiple iterations of "she thinks that he thinks that there are cookies in the jar" (Dennett, 1978; Perner, 1988; Perner & Wimmer, 1985).

The title of this dissertation begins with a recursive iteration "Representing minds representing minds", and so serves as a preamble to the general concept central to the work presented throughout the next chapters. False-belief understanding is indeed useful in a wide array of situations, but the skills that individuals develop when they come to think recursively about beliefs represent another key feature of human sociality that is much less studied empirically (Corballis, 2014; Dennett, 1987; Dunbar, 2011; Kinderman et al., 1998; Miller et al., 1970; Perner & Wimmer, 1985). Recursion, in its simplest conceptualization, is defined as a process that calls itself, allowing for the computation of infinitely long strings of mutually linked elements (Chomsky, 1957; Davis, 1965). Applied to ToM, the principle of recursion is expressed in the form of nested mental states, or mental states about mental states, often termed *Higher-Order* ToM (also variously called *recursive mindreading* or *recursive intentionality*; Dennett, 1978; Miller, 2009, 2012; Perner & Wimmer, 1985). For example, the true state of the world can be considered as zero-order (e.g., the cookie jar is empty), while thinking about an agent's thoughts can be considered as first-order (e.g., she thinks that there is a cookie in the jar), and thinking about an agent's thoughts about another agent's thoughts is equivalent to second-order (e.g., he thinks that she thinks that there is a cookie in the jar), and so on. In this dissertation we are particularly interested in *second-order* ToM, because it represents the first level at which recursion is involved in the attribution of mental states to other individuals (Dennett, 1978; Perner, 1988).

Research in this domain thus far has revolved around an important, but rather narrow range of questions. For instance, we know that the acquisition of second-order ToM is part of children's normative development, and that even though taxing of cognitive resources, it is well within the scope of adults' social reasoning. We also know that adults vary greatly in their ability and propensity to use their ToM skills across all kinds of situations. Yet, as a field, we have very limited knowledge of what cognitive abilities underpin second-order ToM and of what factors affect its emergence and expression. The overarching aim of the work presented in this dissertation is to advance our knowledge of the cognitive processes underlying second-order

ToM. Specifically, the central focus of the dissertation is to examine the role of Executive Functioning (EF) in recursive belief understanding, particularly in adults, although we present some relevant data from children as well.

In brief, EF refers to cognitive abilities that are indispensable for purposeful, goal-directed activity. The period between three and five years of age, which coincides with the development of first-order ToM, is a particularly dynamic period of improvement in EF (e.g., Best & Miller, 2010; Diamond, 2013; Garon et al., 2008). Through the preschool period children come to exert increasing control over their thoughts and actions, which, in turn, promotes adaptive functioning and facilitates the attainment of new and more elaborate goals. Importantly, significant refinement of those skills is also observed through adolescence and early adulthood, often in the form of more efficient information processing, more flexible use of feedback, and more effective decision-making (De Luca et al., 2003; Luciana & Nelson, 2002; Luciana et al., 2004; Luna et al., 2004; Lyons-Warren et al., 2004).

EF has been thoroughly examined in the context of multiple ToM abilities. In particular, a vast literature now attests to lifelong correlations between EF and one of the most well-studied of these abilities, First-Order False-Belief (FOFB) understanding (e.g., Apperly et al., 2009; Bradford et al., 2015; Cane et al., 2017; Carlson & Moses, 2001; Carlson et al., 2002; Carlson et al., 2004; Carlson et al., 2015; Devine & Hughes, 2014; Hala et al., 2003; Frye et al., 1995; German & Hehman, 2006; Hughes, 1998; Hughes & Ensor, 2007; Sabbagh et al., 2006). Although there is a general assumption of cognitive continuity through more advanced forms of belief reasoning, whether the impact of EF is largely confined to the early developmental years, or is instead still relevant for the emergence or maintenance of more advanced skills later in life has received very little empirical attention. In adults, specifically, this question remains largely unanswered. To the best of our knowledge, this thesis represents the first in-depth review of second-order ToM and EF in adults.

The research presented here is organized as follows: Chapters 2 and 3 lay the foundations of the dissertation, providing relevant theoretical background and a review of empirical evidence bearing on questions about the cognitive basis of ToM and EF. Chapter 4 presents Study 1, in which we summarize prior research findings on the association between second-order ToM and EF in both children and adults using meta-analytic methods. Chapter 5 presents Study 2, where we conducted a comprehensive assessment of associations between second-order ToM and EF in a sample of 100 normally-developed adults. In Chapter 6, we report the findings of Study 3, which addresses the relation between second-order ToM and EF with increased specificity in a different sample of 120 normally-developed adults. Lastly, the closing chapter of this dissertation provides a summary of the major findings and some final thoughts on potential ways forward in exploring one of humans' most advanced forms of social cognition.

CHAPTER 2

Recursive Belief Attribution: Developmental Origins

This dissertation focuses in large part on second-order mental states, though first-order mental states will evidently be utilized as reference points along the way. The emphasis on FOFB as a key construct of human sociality has such a longstanding history, that the literature it has generated is immensely useful in informing the workings of further advances in ToM. Therefore, before discussing the particular case of recursion in belief attribution, we review some of the key concepts and early developments that are most theoretically relevant to the work presented in this thesis.

How do Humans come to Think about Beliefs?

ToM owes its research debut to a landmark study by primatologists Premack and Woodruff (1978; Woodruff & Premack, 1979), who investigated the possibility that some non-human primate species could be aware of the subjective nature of the mind. Their work, though not definitively answering that question, inspired what is now widely known as the classic "unexpected-transfer false-belief task" (Baron-Cohen et al., 1985; Wimmer & Perner, 1983). In that task, an agent puts a desirable object in location A before exiting the scene. During her absence, the object is transferred from location A to a nearby location B. Upon her return, participants are asked to indicate where, of location A or location B, the agent will go to retrieve the object.

A meta-analysis of data collected from over 500 false-belief conditions across several different countries revealed that, despite variations in task format, administration, and content, children's performance consistently transitions from below chance level at age three, to above chance around the age of four or five (Wellman et al., 2001). When asked to make explicit predictions about the actions of a given false-belief holder, the majority of 3-year-olds provide realist responses to the test question (e.g., “She thinks that the object is in location B (its actual location)”). A different pattern of responses emerges a year or two later, as children progressively come to see others’ minds as subjective entities, and consider that the content of others’ beliefs does not always match reality. This understanding is believed to consolidate around children’s fifth birthday, the age at which they reliably answer false-belief questions (e.g., “She thinks that the object is in location A (the place the agent last saw it)”).

Interpreting the transition from false-belief failure to false-belief success as a “shift” or a “revolution” in children’s conceptual understanding of the mind is perhaps the most common view regarding the acquisition of false-belief understanding (e.g., Butterfill & Apperly, 2013; Gopnik & Wellman, 1994; Low & Perner, 2012; Perner, 1991; Ruffman, 2014). However, a wide range of divergent theoretical contentions exist. One such contention places a much stronger emphasis on children’s potentially innate - or early-developing - ToM abilities. According to this view, children who fail at identifying false-beliefs may be overwhelmed by the extraneous difficulty of the task itself - e.g., sustaining attention through long story narratives, holding in mind a number of different pieces of information, verbalizing or selecting an appropriate response, etc. Their failure may thus represent a processing deficit rather than a conceptual deficit as such (Leslie et al., 2004; Meltzoff, 2011; Onishi & Baillargeon, 2005; Scott & Baillargeon, 2017).

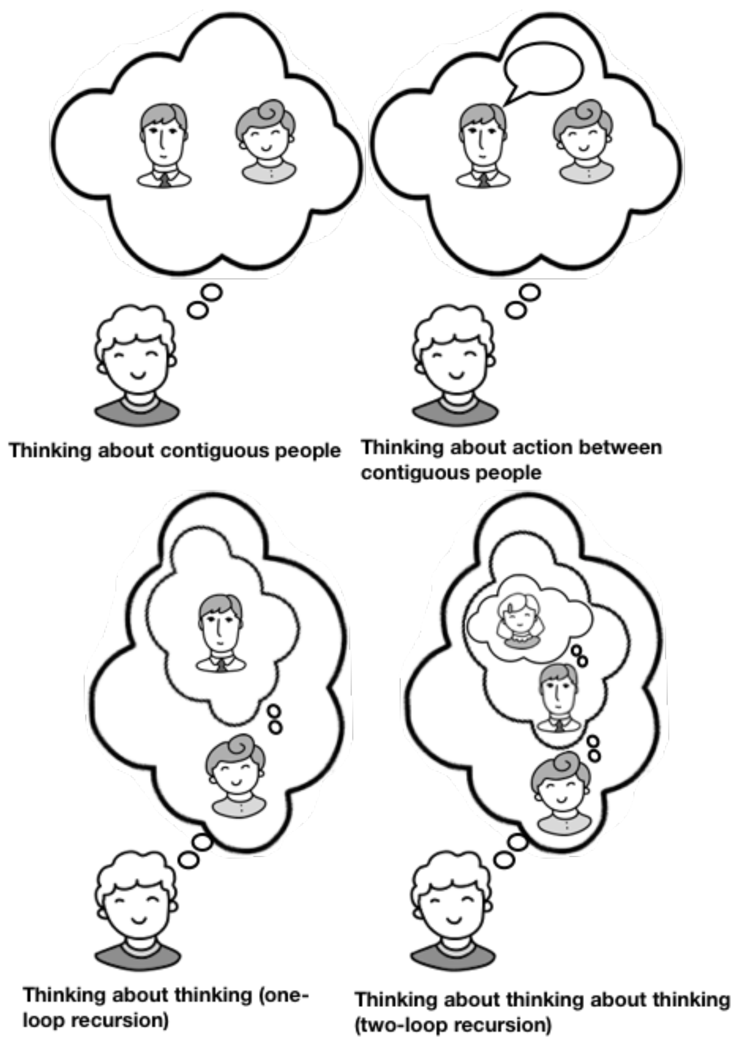
The acquisition of false-belief understanding is indeed a crucial milestone in ToM development, but it does not by any means signal the end of that development. ToM continues to grow in sophistication beyond the preschool years, giving rise to new and more advanced skills for social cognition. There is a (comparatively much smaller) body of research interested in those post-false-belief advances, mainly carried out in the domain of second-order ToM and recursive thinking more broadly. In the following subsection, we turn to this research and review the normative developmental time course associated with second-order ToM along with some of its precursors.

The Development of Recursive Thinking

The recursive nature of ToM finds its experimental origins in studies looking at the broader concept of recursive thinking. The recursive thinking approach was most famously initiated by Miller, Kessell and Flavell (1970), who were interested in uncovering how children came to represent different kinds of hierarchical structures, such as embedded sequences of thoughts about thoughts. The authors reasoned that mental representations, in comparison to motor actions for example, must be fundamentally more complex as they afford the possibility to take themselves as their own object of interest (Miller et al., 1970; Oppenheimer, 1978; Wellman et al., 1996). One can form long series of self-embedded thoughts, such as "I think that you think that he thinks that…", but cannot form (non-representational) action sequences with such levels of dependencies. Inferring recursive thoughts, then, must be more cognitively demanding than processing other, non-recursive, sequences. To test this hypothesis, the authors presented 6- to 13-year-old children with four different types of thought structures conveyed via cartoon drawings. Different sets of drawings depicted a boy with thought clouds or speech bubbles over his head, the content of which illustrated either: (a) contiguity thinking (i.e., thinking about one or more non-interacting persons), (b) thinking about actions between people (i.e. thinking about someone talking to another person), (c) one-loop recursive thinking (i.e. thinking about someone who is thinking about someone), or (d) two-loop recursive thinking (i.e., thinking about someone who is thinking about someone who is thinking about someone). The child participant was asked to verbally describe the illustrations, by responding to a "what is the boy thinking?" prompt. This task is depicted in Figure 1.

Figure 1

Examples of Items used in Miller et al.'s (1970) Recursive Thinking Study



Note. This figure was adapted from: Miller, P. H., Kessel, F. S., & Flavell, J. H. (1970). Thinking about people thinking about people thinking about...: A study of social cognitive development. *Child Development*, 613-616.

Miller et al. (1970) found a stepwise progression in children's understanding of these thought structures. Non-recursive thinking (i.e., contiguity and action thinking) was correctly described much earlier in childhood than one-loop recursive thinking, followed by two-loop recursive thinking. Since the original study, similar developmental patterns have been found in several studies of the same nature (Eliot et al., 1979; Müeller & Overton, 2010; Oppenheimer, 1986; van den Bos et al., 2016), and have been taken as evidence that recursive thinking is slow-developing and perhaps qualitatively distinct from other types of thought processes. Nonetheless, the older children were able to verbalize quite advanced conceptualizations of other's thoughts, as indicated by their successful performance on the two-loop recursion illustrations.

Recursive Thinking Beyond Childhood

The study of recursive thinking beyond childhood has interestingly been much more multidisciplinary than its developmental counterpart, being driven in large part by evolutionary questions. Given how seemingly unusual it is for a species to be able to work at such high levels of representation, researchers were interested in uncovering, for example, how many levels of recursive mental states the human mind could handle without failure, and how their abilities in this regard compared with those of other non-human animals (e.g., Devaine et al., 2017; Kinderman et al., 1998). This approach has motivated an entire literature with a research agenda of its own, along with a strong inclination towards testing adults almost exclusively on the highest levels humanly possible (i.e., two and much higher), with much less attention directed to the mechanisms underlying such advanced abilities.

Most work on adult recursive thinking thus far has relied upon a particular paradigm known as the "Imposing Memory task" (IMT; Kinderman et al., 1998). In the IMT, participants listen to a set of narrated vignettes, each depicting a different social scenario conducive to mental state inferences (e.g., a group of friends organizing a surprise party). The scenarios vary in complexity; some only include two or three characters whose (embedded) mental states must be tracked and later recalled, while others depict more complex situations involving up to seven different characters (e.g., Sophie wants John to think that Sandra wants to organize a surprise party for Stephen). Each vignette is followed by a series of true/false questions, assessing participants' comprehension of the story along progressively increasing orders of recursion. Sometimes participants are also asked to complete "factual" questions, which essentially consist of recalling the sequence of events and so serve as memory checks (Lewis et al., 2011; Stiller & Dunbar, 2007). A "mentalizing" score is then calculated for each participant, on the basis of the longest mental state chain correctly reported. Examples of mentalizing and memory questions presented to participants are listed below (Figure 2).

Figure 2

Examples of Recursive Mentalizing and Memory Questions featured in the IMT

	Mentalizing Questions	**Answer**	**Level**
1	Sam wanted to buy a stamp	True/ False	2
2	Henry thought Sam knew he was a prankster	True/False	3
3	Henry knew Sam believed he knew where the Post Office was	True/False	4
4	Sam thought Henry knew the Post Office was in Bold Street and hence that Henry must have intended to mislead Sam	True/False	5
5	Pete wanted Sam to know that Henry believed that the Post Office was on Elm Street and hence did not intend to mislead him	True/False	6

Memory Questions	Answer	Level
1 Sam needed a Tax Disc from the Post Office	True/ False	1
2 The Post Office was closed because it had moved to Bold Street	True/False	2
3 Sam left Bold Street, then went to the office and spoke to Pete	True/False	3
4 Pete, the man who worked at the same place as Henry, and who knew that Henry was the office prankster, was Sam's cousin	True/False	4
5 Henry, the man that Sam spoke to about where to buy a Tax Disk after he realized he needed to buy one soon, was a colleague of Peter's.	True/False	4

Note. This figure was adapted from Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89*(2), 191-204.

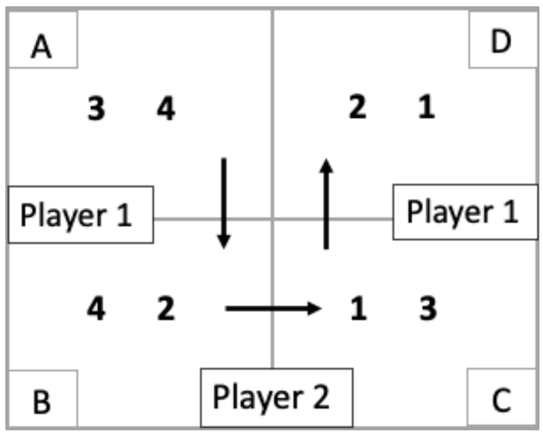
Although it does not address the capacity to *infer* iterated mental states, the IMT is informative to the extent that it provides an estimate of humans' cognitive limits within the domain of socially recursive attributions. In general terms, adults perform optimally as long as the number of mental states to be inferred does not exceed five or six, although some process significantly more or less (Kinderman et al., 1998; Lewis et al., 2017; O'Grady et al., 2015; Powell et al., 2010). From a developmental standpoint, mature competence develops only gradually during childhood, and it is not until late adolescence that adult-like capacity is observed (Dumontheil et al., 2010; Liddle & Nettle, 2006; Valle et al., 2015). Even then, there is evidence that thinking recursively about mental states (e.g., Emma thinks that John wants Emma's boss to give her a pay rise), remains significantly more effortful for adults than thinking about factual statements of equal length (e.g., Emma went to see John about her boss giving her a pay rise) (e.g., Lewis et al., 2017).

The questions concerning the evolutionary motives through which humans might have developed recursive skills has also generated a small body of literature at the intersection of ToM and Game Theory. Cooperative and competitive interactions, for instance, are two broad classes of social activities presumed to provide humans with an adaptive advantage over other species. It may be, on the one hand, that higher-order ToM allows individuals to deceive and manipulate others more effectively by thinking one step ahead of their competitors (Byrne & Whiten, 1988; Devaine et al., 2014; de Weerd et al., 2013; Goodie et al., 2012; Santos et al., 2007). On the other hand, higher-order ToM may be a function of cooperative tendencies, where humans achieve better outcomes by thinking "together" and mutually believing that the other believes that they intend to cooperate (Barrett et al., 2010; Clark, 1996; Lewis, 1969; Schelling, 1960; Scott-Phillips, 2015; Searle, 1995; Tomasello, 2009; Vygotsky, 1980).

A number of strategic games have been devised in order to tease these two hypotheses apart. Although the debate is still ongoing, these strategic game paradigms are relevant to the purpose of this dissertation because they involve recursive thinking and generally reveal great variability across individuals. A widely used competitive task, for example, involves a sequential-move game (typically called "matching pennies" or "hide and seek") in which players have to reason about one another's moves, because a player's payoffs depend on what the other players do, and vice versa (see Figure 3). Typically, the game is played on a 2x2 matrix, with each cell containing separate payoffs for each of the two players. The game starts in cell A, and players take turns at deciding whether to stay in one cell, or to continue on to the next cell. In each cell, the first payoff (i.e., "3") is player 1's, and the second payoff (i.e., "4") is player 2's. If a player decides to stay in a particular cell, the game ends and both players obtain the respective payoffs in that cell. If a player decides to move on, the turn passes to the other player. Players thus maximize their payoffs by using a second-order strategy, whereby one tries to anticipate what the other player might think that they, themselves, think (Goodie et al., 2012; Meijering et al., 2011). For example, in Figure 3, if Player 1 goes first, their optimal strategy to maximize their own gains is to continue on to cell B, because they think that Player 2 will decide to stop the game in cell B, on the basis that they think Player 1 will decide to continue from cell C to cell D. In this example, the recursive structure is thus that “Player 1 thinks that Player 2 thinks that Player 1 will do X because Y”.

Figure 3

Example of a Competitive Game featured in Meijering et al. (2011)



Note. This figure was adapted from: Meijering, B., Van Rijn, H., Taatgen, N., & Verbrugge, R. (2011). I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Participants’ depth of reasoning in these types of paradigms is not assessed directly, but is assumed to correspond to their overall payoffs. That is, a player who engages in more sophisticated thinking will typically achieve larger payoffs than players who choose less sophisticated strategies. Across studies, the most common finding is that participants’ depth of reasoning is significantly prone to individual differences. Only about 3 in 4 adults adopt an optimal second-order strategy to maximize their own gains, while a substantial proportion of participants engage in a first-order strategy, and some participants even play nonstrategically (equivalent to zero-order strategy) (Burchardi & Penczynski, 2014; Camerer, 2003; Camerer et al., 2015; Flobbe et al., 2008; Goodie et al., 2012; Meijering et al., 2011).

The particular factors that account for these individual differences are largely unknown, as the focus of these studies is often on strict behaviour analysis rather than underlying cognitive mechanisms. There is, however, an assumption that the reason why only some individuals are willing to think more deeply about other players' strategies is the cognitive cost associated with such deep thinking. As humans we all have bounded rationality, in the sense that unlike computers, we can only think through so many iterations of the same problem. But for some individuals, it may be too costly to even think through one or two of these iterations (Hedden & Zhang, 2002; Hollebrandse et al., 2008; Meijering et al., 2010; Meijering et al., 2013; Verbrugge, 2009). Why that is so remains to be clarified.

The Particular Case of Recursion in Belief Attribution

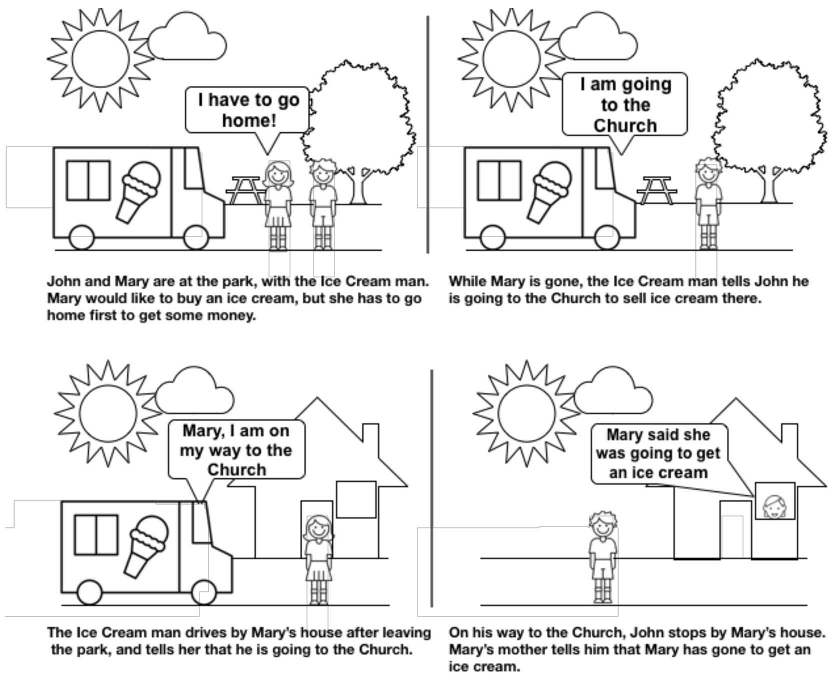
Second-Order False-Belief Understanding in Children

The studies presented in the previous section are informative in showing that from a relatively young age, individuals can comprehend and describe different types of thought structures, including recursive mental states, but that there is nonetheless considerable variability even in adults' performance. The hallmark of acquiring second-order ToM, however, is the ability to ascribe recursive *false-beliefs*, which was not formally studied until Perner and Wimmer (1985) introduced their Second-Order False-Belief (SOFB) task.

Perner and Wimmer (1985) devised a methodology comparable to standard FOFB tasks, with the exception that the false-belief held by the main character concerns the beliefs of a second character rather than the location of an object. Their seminal "Ice Cream Truck" story was originally presented according to the following sequence: Two characters, John and Mary, are in one location, the park, while the ice cream man drives by. Mary would like to buy some ice cream, but she has no money, so the ice cream man says to Mary "you can go home to get your money, I'll be in the park all day". However, after her departure the ice cream man tells John that he is driving over to the church to sell ice cream there. On his way, the ice cream man drives by Mary's house and informs her that he is going to the church. A little later in the day, John goes to Mary's house to ask for help with his homework. Mary's mother answers the door and says: "Mary just left, she said she was going to buy an ice cream". The child participant is then asked the test question: "Where does John think Mary thinks that the ice cream man is?". Individuals who succeed at this task correctly indicate that John believes that Mary believes that the ice cream man is at the park, despite the fact that Mary does not actually hold a false-belief about the ice cream man's whereabouts. This task is depicted in Figure 4.

Figure 4

Example of the Ice Cream Truck story featured in Perner and Wimmer (1985)



Note. This figure was adapted from: Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to-10-year-old children. *Journal of experimental child psychology*, *39*(3), 437-471.

This prototypical task illustrates the three fundamental steps required to induce a SOFB:

(1) Both characters A (Mary) and B (John) have info X (van at the park)

(2) Unbeknownst to A, B gets the information that X changes to Y (van at the church)

(3) Unbeknownst to B, A also gets the information that X changes to Y (van at the church)

For a child, the additional step 3 following the attribution of a FOFB (as in step 2) poses an important computational challenge: Typically-developing children do not begin to show SOFB competence until 5.5-6 years of age. Hence, once children have FOFB understanding, it takes them an average of a further 1.5 to 2 years to be able to integrate beliefs within other beliefs, and to finally comprehend the level of abstraction inherent to the SOFB task. There is thus a demonstrable lag between the initial acquisition of false-belief concepts, and the development of ensuing recursive false-belief concepts.

Following its publication over 35 years ago, there have been attempts at simplifying the design of the Perner and Wimmer (1985) task to examine whether children's difficulty is influenced by incidental demands. These attempts have included providing children with memory aids (e.g., "Remember, John does not know that Mary knows that the ice cream man went to the school"), corrective feedback (e.g., John thinks that Mary thinks that the ice cream man is still at the park), or ignorance questions (e.g., "Does John know that Mary knows where the ice cream man is?") (Astington et al., 2002; Coull et al., 2006; Sullivan et al., 1994). Although data from these modified tasks collectively suggest that the age of mastery indicated by Perner and Wimmer (1985) might have been slightly overestimated, there is currently no available evidence of substantially earlier onset (see Miller, 2009, for a systematic review).

Two main contentions have been proposed to explain the lag between the acquisition of first-order and second-order ToM. The *conceptual change* account considers the development of second-order ToM as arising from the acquisition of new concepts by children over the year or two following the emergence of first-order ToM. According to Miller's (2009) formulation, these conceptual advances involve the idea that beliefs "can have other beliefs and not just events in the world as their target; that such beliefs, like beliefs about the world, can sometimes be false;

and that beliefs about beliefs can enter into recursive chains of potentially any length" (p. 751). Grasping the idea that others can also entertain thoughts about one's own or other people's ToM, or monitoring the relations among multiple mental states and how they interact with one another, may require new ways of thinking about the mind for which younger children may not yet have the cognitive maturity (Perner, 1988; Perner & Wimmer, 1985).

An alternative *complexity account* posits that the classic SOFB task is challenging for children not because of the nature of the mental state inferences per se, but because of the extraneous processing demands imposed by the task. This account implies that if one can anticipate the actions of an agent holding a false-belief, then one should also be able to reason about the consequences of holding a false-belief with any representational content. Whether beliefs are about an object's location, or about another belief, the child who has already learned to recognize others' unseen mental states would only need to make progress in domain general capacities, such as language or EF, in order to pass more difficult tasks such as SOFB (Apperly, 2012; Baron-Cohen et al., 1997; Leslie et al., 2005; Sullivan et al., 1994).

Second-Order False-Belief Reasoning in Adults

Perhaps surprisingly, the ability to ascribe SOFBs is the least studied component of recursive thinking in adults. SOFB reasoning has been traditionally seen as an "all or nothing" ability, meaning that if children from approximately 5-6 years of age perform well on such tasks like the Ice Cream Truck scenario, then older individuals must also have the necessary knowledge to think through the recursive beliefs of others. By the age of 8-9 years, individuals perform at ceiling on these tasks, and because of the dichotomous pass/fail nature of the tasks, age improvements are virtually impossible to detect at later ages. Accordingly, there has been little attempt to explore nuances in adults' SOFB understanding.

One exception to that comes from research by Valle et al. (2015) who devised a third-order FB task with the goal of examining the progression of recursive ToM through adolescence and young adulthood. The task followed the general logic of the Ice Cream Truck task, depicting three brothers (Mark, James, and Luke) and the unexpected transfer of a ball. In one of the stories, Mark is playing with a ball in the bedroom, but then decides to go to the kitchen to get a snack. James and Luke see Mark put the ball in a closed box before leaving. While Mark is in the kitchen, James takes the ball and puts it in the closet. James then goes to the bathroom and, while he is gone, Luke moves the ball to the toy chest. On his way to the bathroom, James meets Mark in the hallway and tells him that the ball is now in the closet. Mark returns to the bedroom before James, and Luke tells him that the ball is in the toy chest. Participants are then asked the following questions:

1. Second-order FB question: Where does Luke think that James thinks that the ball is? (answer: the closet)
2. Third-order FB question: Where does Luke think that James thinks that Mark thinks that the ball is? (answer: the closed box)
3. Third-order FB question: Where does James think that Luke thinks that Mark thinks that the ball is? (answer: the closed box)

While there was no effect of age on SOFB performance, age-related improvements emerged on the third-order FB questions. Adults (*M* = 22.8 years) performed significantly better than both younger (*M* = 14.8 years) and older adolescents (*M* = 17.8 years). The authors concluded that recursive ToM continues to grow through adulthood, and that SOFB is understood significantly better than subsequent higher-orders.

Potential Factors Influencing Recursive Thinking and SOFB Understanding

Recursive thinking is a multifaceted and complex skill, likely to depend on many factors. In general terms, there are at least three broad areas of cognition believed to be related to individual differences in recursive thinking: sociality, language, and processing factors.

First, participating in social interactions may promote the ability to think about the thoughts of others and to coordinate multiple viewpoints. If so, the extent of one's recursive mentalizing capacity should correlate strongly with other aptitudes in the social domain. Indeed, individuals who demonstrate higher mentalizing abilities (as assessed by the IMT) appear to be particularly skillful at maintaining multiple close relationships (Powell et al., 2012; Powell et al., 2014; Stiller & Dunbar, 2007). These individuals also exhibit more cooperative (Paal & Bereczkei, 2006) and empathic tendencies towards others (Čavojova et al., 2011; Launay et al., 2015) in addition to scoring higher on personality dimensions associated with agreeableness (Allen et al., 2017; Ferguson & Austin, 2010; Nettle & Liddle, 2008).

In developmental research, relations have also been found between recursive thinking and advanced aspects of children's everyday social competence, such as faux pas recognition (Banerjee et al., 2011), irony understanding (Filippova & Astington, 2008; Happé, 1993; Massaro et al., 2013), figurative language understanding (Caillies & Le Sourn-Bissaoui, 2008) and reasoning about evidence (Astington et al., 2002). Further, some have begun investigating the role of younger siblings (Paine et al., 2020) as sources of variability in children's second-order ToM attributions. The direction of causality between these variables, however, remains to be investigated.

Second, language is often seen as an important tool for the acquisition of first-order ToM (see Milligan et al., 2007, for a review). Developmental relations between FOFB understanding and mastery of grammar and general syntax, for instance, are often found in children around the age of 5 (De Villiers, 2007). In the same way, some aspects of language may help individuals navigate the demands of representing recursive mental states. The tasks involve complex linguistic structures, and typically conclude with a test question involving at least one recursive clause, the understanding of which may draw on a number of language abilities. For instance, there is a possible role for individual differences in syntactic recursion in the SOFB skills of school-age children (Arslan et al., 2017; De Villiers et al., 2014; Polyanskaya et al., 2018). Syntactic recursion is defined as the embedding of a constituent inside a constituent of the same category (Pinker & Jackendoff, 2005), and is often assessed via the comprehension of relative clauses (e.g., Peter read the book *that* his mother gave him). For example, Arslan et al. (2017), presented 6- to 10-year-old children with simple illustrations of animals performing different actions (e.g., a sheep pushing a monkey, or a sheep pushing another sheep pushing a monkey, etc.), followed by questions assessing their recursive syntax comprehension (e.g., in which picture is there a sheep that is pushing a monkey that is pushing a sheep?). They found a significant correlation between children's scores on this task and performance on a SOFB task. While some theorists argue that recursive mentalizing may have contributed to the evolution of recursive syntax, rather than vice versa (Cheney & Seyfarth, 2008; Corballis, 2014; Dunbar, 1998, 2009; Oesch & Dunbar, 2017; Scott-Phillips, 2015; Sperber, 1994, 2000), the evidence so far tends to favour a bidirectional relation between language and ToM (Milligan et al., 2007).

Third, individual differences in second-order ToM may also be influenced by certain cognitive processing factors. Methodologically, most SOFB tasks are quite complex: In addition to the language requirements just discussed, these tasks contain several pieces of information, as well as several different characters, and are depicted via long and elaborate scenarios. A certain level of processing capacity may be necessary to successfully integrate the task information. For instance, in the Arslan et al. (2017) study just presented, the authors also included a working memory task into their design. Interestingly, performance on this task predicted children's SOFB performance above and beyond syntactic recursion. Cognitive factors may thus also contribute to explaining the heterogeneity often found across individuals' skills for recursive thinking. More precisely, EF is one of the most robust correlates of first-order ToM, and may well be involved in its recursive counterpart. Throughout the remainder of this dissertation, we focus on exploring the relations between EF and ToM, and make the case for continued covariance of these constructs beyond childhood.

CHAPTER 3

Belief Reasoning: What are the Executive Correlates?

In Chapter 2, we presented a brief synthesis of the ToM literature, including some of the debates regarding the nature of the cognitive skills that children display in the early years of life, as well as some further milestones that all typically-developing individuals are expected to achieve once they master the foundational concept of false-belief. We also reviewed the findings of prior work suggesting that focusing on the higher-order mental states of others is prone to individual differences, and that under certain circumstances, adults differ in their propensity to use their ToM recursively. This third chapter turns to one of the most robustly established correlates of false-belief understanding: EF. We begin by outlining the general framework under which EF is best understood, then summarize the literature weaving together EF and ToM.

Executive Functioning: A General Framework of Three Components

EF involves the voluntary control and coordination of cognitive operations, and occupies a central role in a number of areas across the psychological and neurocognitive sciences (Diamond, 2013; Hughes, 2002; Miyake et al., 2000b; Zelazo et al., 1997). When introducing ToM, authors typically begin with a definition of false-belief understanding followed by a walk-through of the gold standard "unexpected transfer" test. In contrast, for EF researchers, there is no such benchmark ability or acid test on which to rely. The cognitive abilities encompassed by the construct of EF are multiple and heterogeneous, making the task of assessing it considerably complex. Further, there are, by most accounts, at least three related but dissociable components that make up the construct of EF: (1) inhibitory control, (2) working memory, and (3) cognitive flexibility. Throughout this dissertation, we will generally follow Miyake et al.'s taxonomy

(2000b), suggesting that each of these three EF components also comprise its own sets of skills, sub-components, and accompanying methods of assessment.

Inhibitory control is the component of EF dealing with the suppression of natural impulses, distracting stimuli, or salient information (Roberts et al., 1998; Rothbart & Posner, 1985; Stuss & Benson, 1986). Different sub-components of inhibitory control include cognitive inhibition and motor inhibition. Cognitive inhibition is defined as the capacity to control interference from external (environmental) or internal (e.g., intrusive thoughts) stimuli, which helps maintain attention focussed on a task in spite of distractors (Harnishfeger & Bjorklund, 1994). These abilities are often assessed via tasks in which particular stimulus-response contingencies must be inhibited (Eriksen & Eriksen, 1974; Simon, 1969; Stroop, 1935). In the Simon task, for instance, participants are instructed to respond to a "blue" circle by pressing the left key, and to a "red" circle by pressing the right key. Cognitive inhibition is triggered on trials where there is incompatibility between stimulus position and response mapping (e.g., red circle presented on the left-hand side of the screen). The incompatibility typically causes interference and slower reaction times (Fan et al., 2002; Kornblum et al., 1990; Lu & Proctor, 1995; MacLeod, 1991). Motor control is generally viewed as the capacity to deliberately withhold a prepotent or ongoing motor response (Evenden, 1999; Logan & Cowan 1984). Motor control assessments are typically based on Go/No-Go paradigms, such as stop-signal tasks (Perugini et al., 2000; van Mourik et al., 2005; Verbruggen & Logan, 2008). In the stop-signal paradigm, participants perform a simple "go" task such as reporting the identity of a stimulus (e.g., a "square" requires a left-key response and a "circle" requires a right-key response). On one-fourth of the trials, the go stimulus is followed by an auditory stop signal, prompting participants to withhold their response on that trial (Verbruggen & Logan, 2008). It is believed that successfully withholding a motor response requires the interaction of a fast control mechanism preventing the execution of the response, and a slower control mechanism monitoring and adjusting performance (Logan & Cowan, 1984).

Working memory is a type of temporary attentional capacity for maintenance and representation of information in mind (Baddeley & Hitch, 1974; Danemnan & Carpenter, 1980; Miller et al., 1960). The continuous/dynamic updating of goal-relevant information in working memory is achieved via the coordination of outdated versus new information. That coordination requires one of two processes: solely *storage*, or storage and *manipulation* of this new information (Conway et al., 2003; D'esposito et al., 1999; Engle et al., 1999; Smith & Jonides, 1999). Storage-related tasks are designed to isolate one's item-maintenance capacity from other working memory processes. In popular span procedures, for example, the number of items to be recalled (e.g., single digits, words, letters) is progressively increased over successive trials, until performance falls below a given criterion level of accuracy. Manipulation-related tasks, on the other hand, are believed to involve some type of memory span procedure, but with added information-processing constraints (e.g., recalling a series of memorized digits, but in the reverse order as presented; see Conway et al., 2005 for a review).

Cognitive flexibility encompasses the capacity to adapt our mental sets to changes occurring in the environment, and to utilize/act on feedback in order to produce appropriate responses (Dajani & Uddin, 2015; Mayr & Kliegl, 2000). Two main sub-components of cognitive flexibility are task-switching and set-shifting. Task-switching measures are generally based on this structure: Participants are first instructed to perform two tasks (e.g., two simple A-B categorization tasks) separately and in succession (e.g., AAA, BBB). Then, the tasks are presented in an alternating sequence and participants must switch between performing one task or the other (e.g., AABBAABB) (Brass & von Cramon, 2004). Set-shifting tasks have the participant perform the same task a number of times, but under different sets of instructions. In this context, a set can be defined as the property of a given stimulus that is relevant in a given trial (Rushworth et al., 2005). For example, the task may consist in determining whether the number in a number-letter pair is even or odd when the pair is presented in a certain location, or determining whether the letter in the pair is a vowel or consonant when the pair is presented in another location. In both cases, cognitive inflexibility is manifested via perseverative responses, or consistent errors, on a task or set clashing with a previous one. Cognitive costs (in reaction times/error rates) arise from difficulties in refocusing attention and modifying one's response strategy (Mayr & Kliegl, 2000).

Developmental Considerations

The developmental trajectory of these EF abilities is underpinned by the development of the prefrontal cortex, which exhibits an extended maturation compared to other brain regions (e.g., Casey et al., 2005; Gogtay et al., 2004; Müller et al., 2013; Spencer-Smith & Anderson, 2009). While fundamental aspects of EF are already observable within the first year of life, substantial development occurs during the preschool years (Best & Miller, 2010; Diamond, 2002; Garon et al., 2008, Zelazo & Müller, 2002). For example, children increasingly opt for delayed gratification, by postponing immediate satisfaction for the sake of receiving larger rewards in the future (Baumeister & Vohs, 2003; Happaney et al., 2004; Zelazo & Carlson, 2012). Children also demonstrate greater attentional flexibility and ability to deal with change. Instead of repeating the same behaviour, they learn to let go of their first thought or response to make an appropriate correction (Diamond, 2013; Zelazo, 2006; Zelazo et al., 2008). Developmental studies also show a two-fold expansion in the number of elements that children can accurately hold in mind over the early and middle years of childhood (Chelonis et al., 2000; Conklin et al., 2007; Gathercole et al., 2004; Hitch et al., 2001; Kemps et al., 2000).

Further, continued maturation of these skills is seen from childhood through adulthood (e.g., Band et al., 2000; Buttelmann & Karbach, 2017; Davidson et al., 2006; Huizinga et al., 2006; Leon-Carrion et al., 2004; Williams et al., 1999), and greater EF efficiency is associated with a wide range of favourable outcomes throughout the lifespan. For instance, EF helps individuals maintain good physical and mental health (Crescioni et al., 2011; Miller et al., 2011; Riggs et al., 2010; Silk et al., 2003), contributes to academic readiness and achievement (Blair et al., 2002; Borella et al., 2010; Colé et al., 2014; Duncan et al., 2007; Gathercole et al., 2004), and facilitates compliance to social norms (Broidy et al., 2003; Denson et al., 2011; Moffitt et al., 2011; Saarni, 1999; Winstok, 2009) as well as job success (Bailey, 2007; Welsh & Schmitt-Wilson, 2013).

Miyake et al. (2000b) presented evidence indicating that while inhibitory control, working memory, and cognitive flexibility are distinguishable in adults, they are also closely interrelated constructs (see also Friedman et al., 2006). Interestingly, the very composition of EF is believed to evolve with development, from a more unitary, single latent factor in early childhood (Hughes et al., 2010; Miller et al., 2012; Senn et al., 2004; Wiebe et al., 2008), to at least the three discrete factors emerging from analyses undertaken with older children and adults (Brocki & Bohlin, 2004; Davidson et al., 2006; Huizinga et al., 2006; Lehto et al., 2003). Importantly, the different components of EF frequently work together in different combinations. Many EF tasks trigger more than one component of EF which, in the context of ToM, has direct implications. Some of these implications are reviewed below.

Executive Processes Involved in First-order ToM

Developmental Evidence

Frye, Zelazo, and Palfai (1995) were among the first to link the emergence of ToM abilities to EF in young children. Frye et al. employed a version of the Dimensional Change Card Sort (DCCS), a cognitive flexibility task which involves presenting a series of cards that vary along several dimensions, including shape and colour. Children are asked to sort cards by one dimension (e.g., colour), and then switch to sorting cards by another, incompatible dimension (e.g., shape). Frye et al. discovered that children who switched more flexibly between sorting rules also displayed better mental state understanding across three ToM measures (false-belief, appearance-reality, and representational change). They concluded that children's cognitive flexibility skills may assist in the construction of multiple conflicting perspectives, and so help children reason through false-belief and other ToM problems.

These initial findings have been replicated and extended in various ways. Carlson and Moses (2001), for example, tested 3- to 4-year-olds on a number of ToM tasks (e.g., false-belief, appearance-reality, deceptive pointing), and a battery of tests designed to measure conflict inhibition and delay of prepotent impulses - two somewhat distinct executive aspects of inhibitory control. Their findings provided more precise conclusions regarding the nature of the EF-ToM association.

First, false-belief understanding strongly correlated with children's overall EF, and this correlation held up independently of a range of potentially confounding factors (e.g., age, gender, verbal ability, family size). Second, not all EF measures correlated equally with ToM. The association with false-belief was much stronger when the inhibitory demands of a given task involved the suppression of prepotent responses *and* the activation of a contradictory response - which presumably adds some form of working memory load to the task - than when the task primarily consisted of resisting immediate gratification (e.g., peeking at a gift being wrapped). The authors concluded that children's growing ability to deal with conflict in information processing concorded particularly well with the demands of the false-belief task, in which a combination of inhibitory control and working memory seem to be required for successful task completion. In both cases, the child participant is expected to monitor two (conflicting) representations of the same reality, and exert inhibitory control over the selection of the prepotent but inappropriate representation in accordance with the task instructions (for similar findings, see Carlson et al., 2002; Carlson et al., 2004, Fizke et al., 2014; Hala et al., 2003; Hughes, 1998a; Mutter et al., 2006; Perner et al., 2002).

Many laboratories around the globe have since contributed to the study of EF in relation to the origins of ToM. Building on an earlier meta-analysis by Perner and Lang (1999), Devine and Hughes (2014) pooled the results of over 100 studies linking EF and ToM in childhood, including a range of EF and false-belief tasks. They found that, for preschoolers, progress in EF is strongly associated with progress in belief attribution, and remains strong over and above factors such as age, general intelligence, and cultural background (r = .38). Further, longitudinal designs consistently report that children's EF performance at earlier time points reliably predicts their ToM achievements at later time points. The reverse direction (of early ToM predicting later EF), on the other hand, generally yields much weaker associations (e.g., Carlson et al., 2004; Hughes, 1998; Hughes & Ensor, 2005, 2007; Marcovitch et al., 2015; Müller et al., 2012; Pellicano, 2010; Razza & Blair, 2009; Schneider et al., 2005).

Prevailing Theoretical Conjectures

While the strength of the EF-ToM relation is clear, the basis of the relation is open to interpretation. Two main positions have been proposed: the *expression* account and the *emergence* account (Moses, 2001).

On the *expression* account, ToM skills may already be present in young children, but may not be apparent until a sufficient level of EF has developed. The rationale is as follows: children may fail to resolve a false-belief problem not because they lack the necessary knowledge, but because the traditional task imposes executive overload affecting their ability to demonstrate this knowledge. Errors occur because children do not have enough inhibitory control to overcome the demands of the task (Carlson et al., 1998; Leslie & Polizzi, 1998; Russell et al., 1991).

On the alternative *emergence* account, executive processes may provide some of the foundational skills for the acquisition of false-belief understanding in the first place (Moses, 2001). For instance, an immature capacity to hold back their own point of view may impede the recognition that other points of view are possible. With progress in EF, children learn to inhibit the natural salience of their current knowledge, and realize that instead, other alternative perspectives may be considered. Likewise, children may need sufficient working memory to hold in mind different perspectives, before they can even recognize the notion of perspective (Carlson & Moses, 2001; Moses, 2001; Moses & Tahiroglu, 2010; Russell, 1996).

Lifelong Reliance on Executive Control for Belief Attributions

As adults, we are motivated to entertain all kinds of thoughts and beliefs about other individuals, and routinely do so without particular effort. Occasionally, however, we tend to overly rely on the external context, instead of others' state of mind, or we exhibit egocentrism in our evaluation of what information is or isn't available to other individuals (Bernstein et al., 2011; Birch & Bloom, 2004, 2007; Epley et al. 2004; Keysar et al., 2003; Lagattuta et al., 2014; Sommerville et al., 2013). These biases are expressed more subtly in adults, and often corrected quite rapidly, but they do parallel the deficits (in inhibitory control and/or working memory) that younger children exhibit before their ToM fully develops in the preschool years (Epley et al., 2004). For example, in a manner similar to children, adult participants who know where an object is hidden are likely to overestimate the probability that a naïve newcomer would search in that location to retrieve the object (Birch & Bloom, 2007).

Through demonstrations that belief-tracking is not always initiated automatically, the mindreading skills of adults have indeed been shown to depend on some level of executive control and deliberate effort. Apperly, Riggs, Chiavarino and Samson (2006a), for instance, conducted a seminal study showing that when not explicitly prompted to do so, adult participants do not automatically track the beliefs of others. When presented with simple false-belief scenarios, participants who are unexpectedly asked to indicate where the agent believed the object to be located tend to show a significant delay in responding. The authors suggested that this delay is caused by the need to retrieve the necessary facts, and calculate the agent's belief only *after* there is a task requirement to do so, rather than proactively as automaticity would suggest. This finding has been extensively replicated by different research groups (Back & Apperly, 2010; Bardi et al., 2016; Bradford et al., 2015; Nijhof et al., 2016, but see Cohen & German, 2009 for conflicting findings), and considered as clear demonstration of the non-automaticity of belief ascriptions (at least as assessed under these conditions; see Low et al., 2016 for a review of instances in which belief-tracking may occur automatically).

The fact that there is a processing cost associated with the calculation of others' beliefs may reflect the need to deploy executive resources to accomplish such calculations, even in mature ToM. Supporting this idea, German and Hehman (2006) found that even for young adults, inferring combinations of mental states that place relatively high demands on EF (e.g., false-belief plus negative desire) induced significant reaction time costs, compared with combinations that make lower demands on EF (e.g., true-belief plus positive desire). Thus, even though these tasks may be relatively trivial for adults in terms of their ability to make the correct judgments, the efficiency of such judgements as reflected in reaction times appears to be influenced by the executive demands of the task. These findings are consistent with some correlational data, supporting a role for EF in even some of the most basic forms of mental state reasoning. For example, high working memory capacity adolescents (Nilsen & Bacso, 2017) and adults (Lin et al., 2010; Wardlow, 2013) tend to engage more willingly in perspective-taking behaviours (i.e., inferring what others see/know) than low working memory capacity individuals.

In sum, the research reviewed in this section demonstrates that adults, in a manner similar to children, still exert some deliberate effortful control over their belief-tracking activities. In light of this evidence, it would not be unreasonable to expect that the contribution of EF to belief reasoning would extend to later-developing, more complex abilities such as SOFB. Yet, as we will see in the next chapter, the extent and specific nature of these potential relations remain largely unexplored.

Motivation for the Current Studies

In the preceding sections, we summarised what is currently known regarding the association between EF and first-order ToM. Doing so illustrated how central EF is to the processes supporting false-belief reasoning, especially in the early years of life as children are building their understanding of other minds. We also discussed how adults, even though remarkable mindreaders, can be subject to some of the same constraints experienced by children when the salience of their own knowledge stands in the way of their ability to take different perspectives into consideration.

An important issue with respect to EF is not only its contribution to individual differences in ToM, but also its potentially differential role in first and second-order ToM. Examining the degree of continuity/discontinuity between the specific EFs associated with first-order ToM, and those associated with second-order ToM, may help answer some of the open questions regarding the nature of SOFB reasoning.

While there is no doubt that second-order ToM shares at least some of its key representational features with first-order ToM, it may also give rise to new executive requirements in some ways specific to the process of thinking recursively about beliefs. For instance, there may be some degree of uniqueness to the conceptualization of beliefs about beliefs. To process the multiple perspectives involved in recursive ToM, especially strong working memory and cognitive flexibility skills may be necessary. Or different combinations of EFs (e.g., inhibitory control and cognitive flexibility) could be found to work together and explain why some individuals attribute recursive beliefs with more ease than others (Austin et al., 2014; Bock et al., 2015; Farrant et al., 2014; Lagattuta et al., 2015; Miller et al., 2018; Perner et al., 2002).

Alternatively, the way we represent beliefs about beliefs may be quite similar to the way we represent beliefs about objects or aspects of the world. First and second-order ToM may be supported by relatively similar processes, and thereby also relate to much the same EFs (Apperly, 2012; Baron-Cohen et al., 1997; Leslie et al., 2005; Lockl & Schneider, 2007; Sullivan et al., 1994).

Evidently, as both EF and ToM continue to develop beyond childhood, it is necessary to account for changes in these domains when investigating their interrelations in older individuals. It is likely that contributors to ToM may change over time and depend on the developmental stage examined (Im-Bolter et al., 2016; Meinhardt-Injac et al., 2020). For instance, the extent to which first-order ToM relates to adults' EF is relatively well established, but not as specifically nailed down as in early childhood. It is not entirely clear which aspects of EF are most closely related to first-order ToM in adults, and whether these align with the EFs most closely related to children's developing understanding of false-belief. There is thus no guarantee that the specific relations observed in early childhood will also be observed in individuals with more mature EF and ToM skills. We might find that the relation of SOFB to EF is different than the FOFB-EF relation typically observed in preschool for example, but that could simply be because of changes in the EF-ToM relation over time and experience, rather than an actual difference in the EFs recruited by each construct.

On this basis, if we want to assess the correlates of SOFB in adults, and compare them to the correlates of FOFB, doing so requires a thorough assessment of each construct conjointly. Testing both FOFB and SOFB under the same conditions and within the same participants would provide a solid basis to inform the potential similarities and distinctions in their respective relation to EF. In the studies that follow, we proceed to such assessment in two ways: Study 1 aims at providing an initial review of the current state of knowledge regarding the links between

SOFB and EF in both children and adults, while Studies 2 and 3 are dedicated to an extensive mapping of the adult capacity for second-order ToM reasoning in relation to EF.

CHAPTER 4

Study 1

Individual Differences in Second-Order False-Belief Understanding and Executive Processes: A Meta-Analytic Review of Evidence from School-Age Children and Adults

As discussed in previous chapters, a comprehensive literature has demonstrated that even simpler forms of mental state reasoning, such as attending to the beliefs that someone may hold about objects or aspects of a situation, are prone to limitations across the lifespan, and make lasting demands on executive resources (Apperly et al., 2009; Bradford et al., 2015; Converse et al., 2008; Devine & Hughes, 2014; Epley, 2008; Keysar et al., 2000, 2003; Lin et al., 2010; Mckinnon & Moscovitch, 2007). Similarly, in relation to second-order ToM, a claim is often made that its computations must be cognitively costly (e.g., Dunbar, 1998; Lin et al., 2010, but see O'Grady et al., 2015). The fact that there are limits, for instance, on the number of embedded mental states that the human mind can handle (Kinderman et al., 1998), as well as the finding that most adults do not reliably use their second-order ToM even when provided with a tangible incentive to do so (e.g., winning a game; de Weerd et al., 2013), serve as an indication of the effort required by second-order ToM reasoning, that only some individuals are willing to deploy across contexts.

Yet, the extent to which EF plays a supportive role in these inferences is currently unclear. In this first study, we provide a meta-analytic review of the empirical evidence on the relation between SOFB understanding and EF in a wide age range, spanning the middle childhood years to adulthood. Although adult cognition is the primary focus of this dissertation, we include child data in this study for purposes of comparison. We begin by laying out the rationale for linking SOFB understanding to EF, starting with a review of the aspects of the classic SOFB task that are most likely to depend on EF. Next, we introduce the empirical evidence suggesting a link between SOFB and EF in both children and adults. This evidence is then combined in a quantitative meta-analysis.

SOFB Task Analysis: How might EF be Involved?

At least three elements of the SOFB task may be particularly demanding of executive resources. First, SOFB tasks may involve a response-selection process in some ways “doubling” the inhibitory demands of the first-order task. To recap, in the classic SOFB task both Agents 1 and 2 have info X (e.g., that the ice cream man is at the park); then unbeknownst to Agent 2, Agent 1 receives the information that X changes to Y (e.g., that the ice cream man is going to the church); finally, unbeknownst to Agent 1, Agent 2 also gets the information that X changes to Y (i.e., both agents now know that the ice cream man is going to the church). It is important to reiterate that Agent 2 *does not* actually hold a false belief about the location of the ice cream man, and that the false-belief of interest in this task concerns Agent 1’s false-belief about the fact that Agent 2 knows the actual location of the ice cream man. In this way, it may be especially challenging for an individual to overcome the interference between, not only their own knowledge of the true state of affairs and the false-belief of Agent 1, as in a FOFB scenario, but also between *Agent 2’s* knowledge of the true state of affairs and the false-belief of Agent 1 about that knowledge. For a correct judgment, the participant must resolve a twofold conflict of the type “the ice cream man is at the church, and both Agent 1 and Agent 2 know that the ice cream man is at the church, but Agent 1 does not know that Agent 2 knows that the ice cream man is at the church and thinks that Agent 2 thinks that the ice cream man is still at the park”. The “pull of the real" is in this way arguably doubly difficult to resist, because both the participant and Agent 2 know where the ice cream man is, but this knowledge must be inhibited in order to take Agent 1's false-belief into consideration. Considerable inhibitory control may thus be required to overcome the salience of one's own, in addition to someone else's knowledge of reality (Leslie et al., 2004, 2005; Perner et al., 2002).

Second, individual differences in cognitive flexibility may also be particularly predictive of one's ability to handle recursive beliefs. For instance, once they can successfully infer the FOFB of others, individuals must further realize that beliefs can in fact be about other beliefs, which can also be about yet other beliefs, and so *ad infinitum*. To be able to envision the hierarchical sequence that this generates, individuals likely need to focus on multiple perspectives in successive, and concomitant fashion (Bock et al., 2015). Specifically, in classic SOFB tasks, the mental states of Agent 1, at one point in the sequence, must be integrated within the mental states of Agent 2, which means that the participant must be able to reason through the interactions that link these mental states together. Presumably, doing so would involve a necessary back-and-forth between the task of inferring the beliefs of Agent 1, then the task of inferring the beliefs of Agent 2, and finally the task of inferring the beliefs that Agent 1 and Agent 2 have in common. In the context of the example given above, there is an initial need to calculate each Agent's true belief (e.g., Agent 1 and Agent 2 both believe that the ice cream man is at the park), then Agent 1 acquires a new piece of information and her true belief must be updated (e.g., the ice cream man has moved to the church, and Agent 1 knows that Agent 2 does not know that). At this point, Agent 2's true-belief must be momentarily switched to a false-belief (e.g., Agent 2 does not know that the ice cream man is now at the church), until she acquires the same true-belief information as Agent 1 (e.g., Agent 2 also knows that the ice cream man has moved to the church). Here, participants must return to Agent 1, update her belief from true to false, and integrate it within the context of Agent 2's beliefs (e.g., Agent 1 does not know that Agent 2 knows that the ice cream man is at the church, and still believes that Agent 2 believes that the ice cream man is at the park). Given the considerable number of switches induced by the recursive nature of the task, individuals with higher levels of cognitive flexibility may be able to handle SOFB problems with more efficiency than individuals with lower levels (Bock et al., 2015; Miller et al., 2018).

Third, it is difficult to imagine, of course, how this process could be exclusive of working memory. It is, in fact, practically impossible to ascribe SOFBs without working with a relatively large amount of information no longer perceptually present. Classically, SOFB scenarios not only unfold over significantly longer sequences of events, but there is also a significant delay between different beliefs inferred at different time points. In reference to the example used above, we have illustrated how it is not until the very end of the sequence that the pieces of the belief puzzle must be assembled in order to make the correct SOFB judgement. For example, to be able to predict where Agent 1 believes that Agent 2 believes that the ice cream man is, participants must retrieve some of the belief information encoded much earlier in the sequence, at the point where both agents saw the ice cream man at the park, before he moved from park to the church. The retention period separating these two belief inferences may be one aspect of the task particularly taxing of working memory resources. Another challenge could be the memory load that presumably arises at the point of organizing the encoded belief information into a recursive structure (i.e., integrating the belief of Agent 1 with the belief of Agent 2). This procedure, sometimes referred to as "merging" (Chomsky, 1995), involves holding both beliefs in mind for a period of time long enough for them to be evaluated against each other and combined appropriately. As mentioned above, each agent's beliefs are not only inferred and kept in mind separately, but also as a whole in order to represent their interrelations. That is, there is a requirement to not only retain the belief information but also presumably to manipulate this information in working memory. Finally, relative to the first-order case, there is also, of course, an increase in the number of beliefs to be inferred and held in mind, which could on its own represent a source of working memory difficulty (Apperly et al., 2009; Lecce et al., 2017).

Empirical Evidence of Links between Second-Order ToM and EF

As mentioned earlier, the Devine and Hughes (2014) meta-analysis focused primarily on the emergence of FOFB and its relations with preschoolers' EF skills. Nonetheless, the authors also identified a subset of studies quantifying the relations between SOFB and EF. All of these reports sampled children under the age of six, and the majority used a version of the Ice Cream Truck story described earlier to assess SOFB understanding. Together, the pooled effect sizes revealed a moderate association between SOFB and EF (r = .44), which was largely consistent with the outcome of their FOFB-EF analysis (r = .38). However, a potential caveat is that some of the effect sizes included under the second-order category may have actually been obtained from composite scores including both FOFB and SOFB, such that the exact variance attributed to SOFB remains unknown. Since the publication of this meta-analysis, however, new studies have investigated the links between SOFB and EF in early and middle childhood, and will be systematically combined and analyzed here for the first time.

In contrast, research dedicated to the evolution of these links through adulthood remains sparse. To the best of our knowledge, only one study, that of Valle et al. (2015) presented earlier, conducted a correlational analysis of the relations between SOFB and EF in a pure sample of typically-developed adults. Further cases come from studies of patient populations (e.g., adults with depression, schizophrenia, etc.), which included matched control group data from more typical adults. When considered as individual studies, however, the information that can be drawn from these studies is quite limited and does not allow for clear conclusions. By conducting an exhaustive meta-analytic review, including both published and unpublished data, our goal was to provide more definitive evidence than narrative reviews alone can provide.

Study Aims and Predictions

The main objective of the current meta-analysis was to incorporate empirical evidence from school-age children and adults to draw conclusions on the strength, and potential moderating variables, of the relation between SOFB and EF in typically-developing individuals. For instance, some sample characteristics (e.g., age, culture) and methodological factors (e.g., type of EF tasks, size of test battery) may moderate the SOFB-EF relation within each age group. Where there is sufficient data, we review the influence of such factors. The present meta-analysis was conducted according to a preregistered protocol (osf.io/u6ndv), including the following aims and predictions:

(1) The first aim was to systematize the literature on individual differences in EF and SOFB reasoning in order to assess the size of the relation between the two constructs. Considering the robust empirical support from studies of FOFB and EF, and the task analysis just presented, we expected to find an overall significant association between SOFB and EF.

(2) The second aim was to examine the influence of age on the relation between SOFB and EF. Although we predicted a significant overall correlation for children, we did not have a definitive prediction for adults. Rather, we considered two general possibilities. If the SOFB-EF relation found in Devine and Hughes (2014) does indeed exist and extends beyond the early

school years because of continuing executive task demands, then we should find a large degree of consistency across child and adult samples. Alternatively, if progress in EF is only necessary for the *development* of SOFB in childhood, then the relation between the two constructs should be largely confined to the child samples.

(3) The third aim was to explore the specificity of links between SOFB and EF. The task analysis presented earlier provided initial theoretical support implicating a range of different executive abilities. However, we did not have a direct hypothesis as to whether broad or more specific SOFB-EF linkages would emerge from our analysis. On the one hand, there could be some continuity between the EFs most involved in FOFB, and those most involved in SOFB (i.e., working memory and/or inhibitory control). On the other hand, FOFB and SOFB may differentially correlate with EF, such that certain specific EFs, or combinations of EFs, may be most involved in SOFB, but not FOFB, or vice versa.

Methods

Literature Search Strategy

To identify eligible studies we systematically searched seven electronic databases (PsycINFO, PsycARTICLES, ProQuest Dissertations & Theses Global, Open Access Theses and Dissertations, Google Scholar, OSF Preprints, and PsyArXiv), with combinations of the following keywords: (Second-order) or (Recursive) or (Higher-order) AND (Theory of Mind) or (Mindreading) or (Mentalizing) or (False Belief) or (Social cognition) or (Social perception) or (Social reasoning) AND (Executive function) or (Executive control) or (Executive dysfunction) or (Neurocognitive function) or (Neurocognitive capacities) or (Prefrontal cortex) or (Self-control) or (Self-regulation) or (Cognitive control) or (Inhibition) or (Inhibitory control) or

(Working memory) or (Updating) or (Attention) or (Attentional control) or (Planning) or (Shifting) or (Set-Shifting) or (Switching) or (Cognitive flexibility). We further conducted a manual search of the following: reference lists of retrieved articles and related meta-analyses/systematic reviews, several scientific journals (Child Development, Journal of Cognition and Development, Developmental Psychology, Cognitive Development and Topics in Cognitive Science), conference proceedings for the last three meetings of the Society for Research in Child Development, Cognitive Development Society (CDS), and Jean Piaget Society (JPS). The Google Scholar features “Cited by”, “Related articles", and “User profiles" were also examined in an attempt to locate as many pertinent articles as possible. In addition, we posted on relevant listservs requesting relevant papers or data (CDS, JPS), and researchers who were thought likely to have carried out relevant research were contacted for in press/forthcoming journal articles or additional data (published and/or unpublished). We conducted an initial search in June-July 2019, and a second search in March-April 2020.

Inclusion and exclusion criteria

Studies were selected based on the following core criteria:

(1) Reports needed to include original empirical research (review articles, meta-analyses, editorials or commentaries were excluded).

(2) Study design needed to include a sample of typically-developing children, adolescents, or adults. If data examining the relation between SOFB and EF were available for a typically developing control group in the context of a study addressing the relation for a clinical population, data for the control group were included.

(3) Presence of at least one experimental measure of SOFB understanding.

(4) Presence of at least one behavioural measure of EF (i.e., working memory, inhibitory control, cognitive flexibility, planning, attention).

(5) Given the scarcity of longitudinal studies, data from cross-lagged designs that did not report correlations between SOFB and EF skills at any *concurrent* time points were excluded from this meta-analysis.

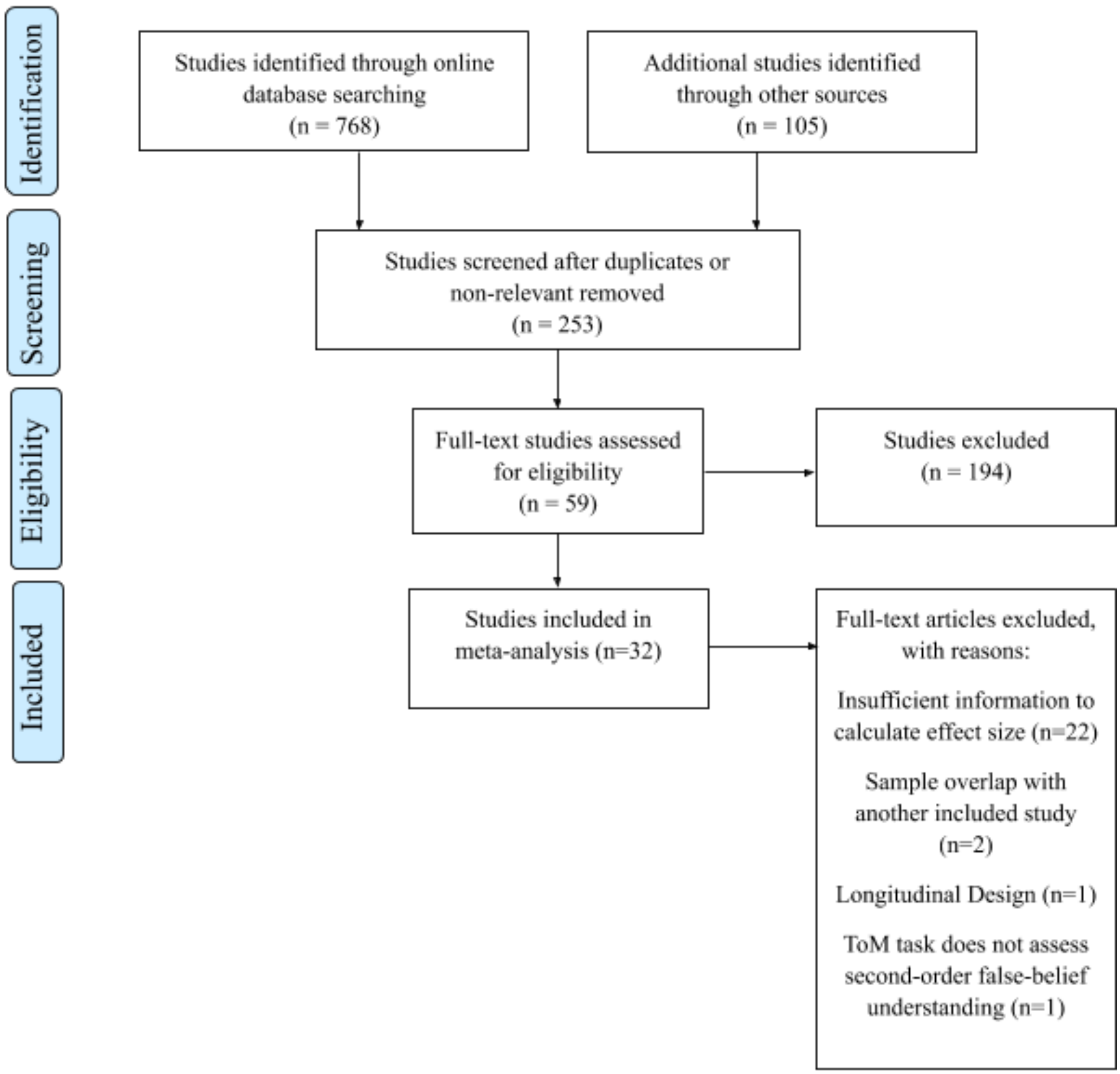
(6) Presence of sufficient statistical information to calculate Pearson's correlations (r) or Spearman's rank-order correlations (rs).

No limits were applied for publication dates, and we considered reports written in either English or French (languages in which members of our team are fluent). If additional data or further specification was needed from a particular eligible study, we contacted the first authors to request the relevant information. A total of forty-nine researchers were contacted; twenty-one authors provided additional data or analyses, thirteen of which met the inclusion criteria and were included in the final dataset. Ten authors no longer had access to the requested data, and eighteen researchers did not respond.

The Systematic Reviews Web Tool Rayyan (Ouzzani et al., 2016) was used to assist in the process of screening and selecting studies, and full texts of all potentially eligible studies were dual-coded in terms of the inclusion criteria by two independent reviewers. Cohen's kappa coefficients indicated an interrater reliability indicative of almost perfect agreement ($\kappa = .95$). The few remaining discrepancies or inconsistencies in data coding were resolved by consensus. Full details of the progression of the studies through exclusion stages is diagrammed in the PRISMA chart in Figure 5 below.

Figure 5

Flowchart of study Selection Procedure, according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)



Note. This diagram was adapted from: Moher, D., Liberati, A., Tetzlaff, J., Altman, DG. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med,6*(7): e1000097.

Meta-analytic procedure

Pearson correlations (*r*) were extracted from each eligible report to assess the strength of the association between SOFB and EF across studies. If a relevant study reported Spearman's *r*s coefficients ($n = 2$), these were treated as equivalent to Pearson's correlations (Myers & Sirois, 2006). All extracted *r* values were then transformed to variance-stabilizing Fisher's *Z*s (Borenstein et al., 2009). However, for ease of interpretation overall effect sizes presented below and in figures were backtransformed to Pearson's *r*. For the majority of extracted effect sizes, higher values indicated better performance. In cases where variables were originally scored in the opposite direction (higher values reflected poorer performance, e.g., reaction time, number of errors), these were reverse coded before analysis, such that all positive values indicated better performance.

Data were analyzed with the meta-analytic Metafor package in *R* (R Core Team, 2019; Viechtbauer, 2010). To compute the main estimates, we used random effects meta-analytic multilevel modelling with the restricted maximum-likelihood method (specified with the rma.mv Metafor function). We began by estimating models accounting for the dependencies between the observations reported by some of the studies included in our database (Cheung, 2014; Konstantopoulos, 2011; Nakagawa & Santos, 2012). For instance, many studies administered multiple EF tests on the same sample of adults or children, and thus reported multiple dependent effect sizes for these constructs. Using a multilevel modelling approach allowed inclusion of all effect sizes in the meta-analytic model, by assigning a study-specific random effect to each effect size observed from the same sample. Then, to evaluate the level of heterogeneity in the effect sizes reported across studies, Higgins' I^2 and Cochran's *Q* statistics were computed with 95% confidence intervals. Given that I^2 values of 25%, 50% and 75% are respectively considered to represent low, moderate and high levels of heterogeneity, an I^2 value above 50%, together with a statistically significant Q statistic are conventionally used as thresholds to indicate that variation in effect sizes across studies may be greater than that expected by chance (Higgins et al., 2003). We additionally applied a cluster-robust estimator (via the *Robust* function in *R* Metafor package; Hedges et al., 2010) to adjust the potentially underestimated standard errors and confidence intervals associated with dependent observations within studies, and implemented a small-sample correction in all models to avoid inflating Type I error (Tipton, 2015).

To attempt to explain heterogeneity in effect sizes, potential moderators were entered separately in the original model as predictors of effect size. Coding of moderators revealed an interrater reliability exceeding a Cohen's kappa of .95. A final analysis included the set of moderators that were viable candidates to explain any variance in the outcome, excluding those with little explanatory power (at $p > 0.05$). Finally, potential publication bias was assessed with a *p*-curve analysis (http://www.p-curve.com/; Simonsohn et al., 2014), along with the following regression-based indices: the funnel plot, Egger's regression test (Egger et al., 1997), and Rosenthal's and Orwin's fail-safe *N* tests (Orwin, 1983).

Results

The selection process resulted in final inclusion of 83 effect sizes (*k*) derived from 32 independent studies (*n*), 30 of which were published in peer-reviewed scientific journals (the remaining effect sizes were obtained from unpublished doctoral dissertations, $n = 2$). The total participants across studies was 2584. The majority of identified effect sizes were extracted from cross-sectional studies ($n = 29$), and our final dataset included samples from 16 different countries. Of the 32 studies retained for analysis, only five examined the relation between SOFB and EF in adults ($k = 11$, $N = 246$). Four of these five studies consisted of investigations with the primary purpose of testing the cognitive abilities of a patient group (e.g., adults with schizophrenia) in relation to various aspects of social functioning. From these studies we extracted control group data ($k = 9$). For children, 27 studies examined the relation between SOFB and EF ($k = 70$, $N = 2338$). Seven of our child samples were also provided by clinical research including a normally developing control group ($k = 25$). Across all studies, sample sizes ranged from 10 to 288 participants, with a mean sample size of 79.91 participants. Child participants were aged 7.57 years on average (range: 5.50-11.96 years), and adult participants had a mean age of 31.71 years (range: 18.46-43.70 years). A table presenting the characteristics of each study included in this meta-analysis can be found in Appendix A.

Adults as well as children were predominantly tested on tasks originally designed to assess children's emerging SOFB reasoning, including Perner and Wimmer's (1985) Ice Cream Truck story ($n = 18$), and Sullivan et al.'s (1994) Birthday puppy ($n = 20$) as the most frequently reported tasks. In regard to EF, working memory was the core executive component most often examined in relation to SOFB performance ($n = 24$, $k = 35$). Inhibitory control was assessed in 16 studies ($k = 20$), and cognitive flexibility was assessed in 14 studies ($k = 19$). Twenty-nine different tasks were employed as indices of executive skills across studies, though the four most commonly reported tasks were consistent across age groups: Dimensional Change Card Sorting task (DCCS), Digit span, Word span, and Stroop Colour/Word task.

We first estimated the overall association between SOFB and EF by fitting an intercept-only model to data from all studies, with random effects allocated to each independent study. The summary effect size estimate was $r = .23$, CI [.17, .29], $p < .001$, $Q(82) = 176.89$, $p < .001$, $I^2 = 58.18\%$, indicating a small overall effect size and moderate-to-high variability in effect sizes across studies (Cohen, 1988; Higgins et al., 2003). Funnel plot analyses showed reasonable symmetry in the reported effect sizes, consistent with Egger's regression test (*p* = .38) suggesting minimal small study bias in the overall dataset (Egger et al., 1997). According to Rosenthal's and Orwin's file-drawer analyses, a large number of additional null effect sizes would be required to reduce the outcome to non-significance (*k* = 8466 for Rosenthal) or below an effect size of .13 (*k* = 83 for Orwin). Finally, our P-curve analysis yielded right-skewed curves, suggesting that the studies included in the meta-analysis contained evidential value, not simply Type 1 errors ($p < .001$).

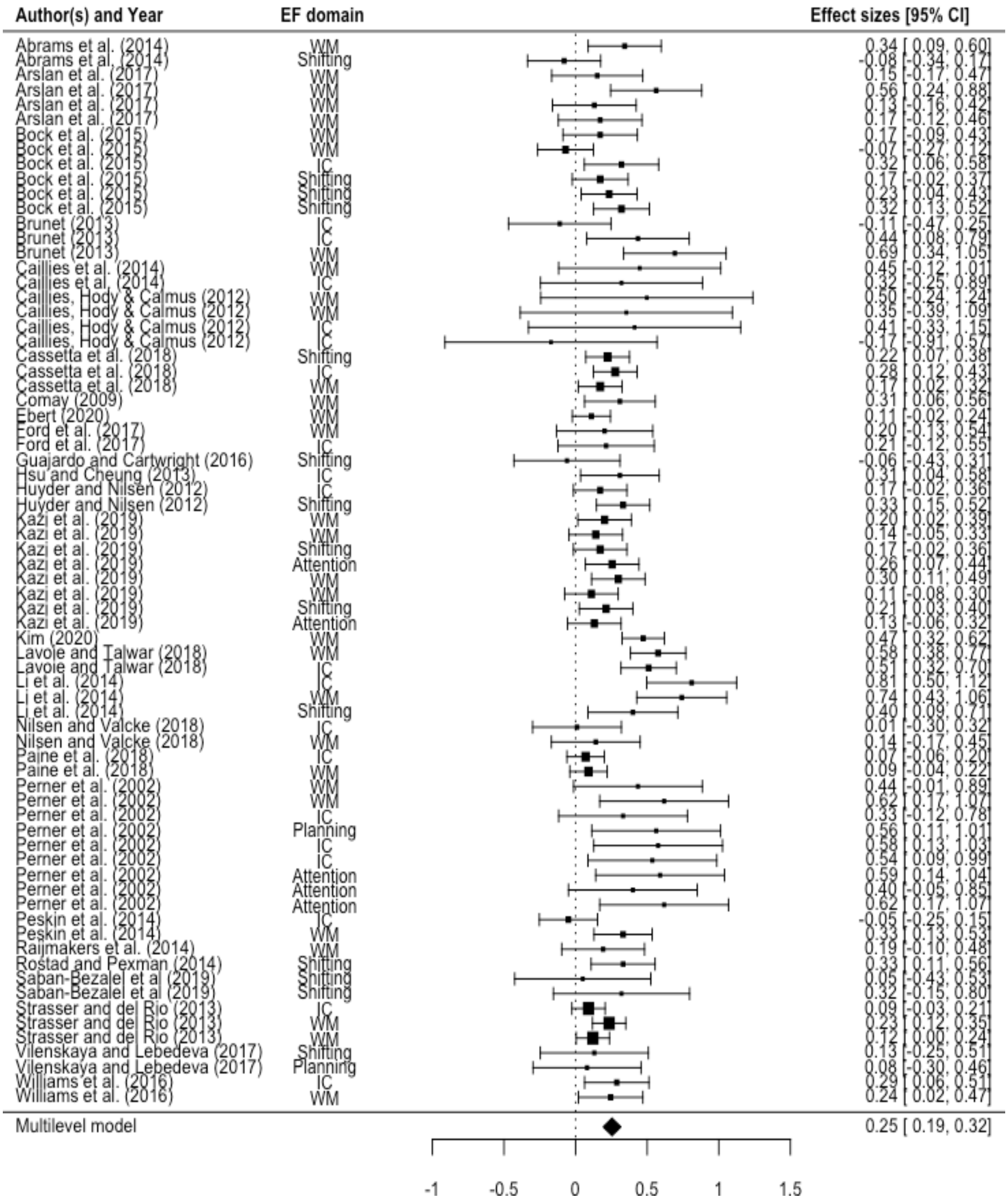
The large heterogeneity in our dataset warranted the inclusion of moderator variables in our model. We extended our intercept-only model to include the age of the participants as a categorical moderating factor (such that adult samples were coded "0", and child samples were coded "1"). This analysis revealed somewhat different overall effect sizes for adult and child samples $F(1,30) = 3.86$, $p = .059$, with a significant amount of heterogeneity within age groups, $Q(81) = 169.78$, $p < .001$, $I^2 = 56.09\%$. In the next section, we present within-subgroup analyses examining the relation between SOFB and EF in child and adult samples separately.

Child samples

Figure 6 below shows a forest plot summarizing the data from 27 studies examining the zero-order relation between children's SOFB and EF. The square markers represent the weighted mean of each individual effect size and the width of the horizontal lines represent the corresponding 95% confidence intervals. The diamond marker at the base of the graph depicts the overall pooled mean effect size, taking account of all 72 effect sizes analyzed for this age group. Effect sizes ranged from $r = -.17$ to $r = .81$, and 35 of them showed a significant positive association between SOFB and EF with none significant in the negative direction. Overall, the relation between SOFB and EF was significant (*r* = .25, CI [.19, .32], $p < .001$), with considerable remaining heterogeneity, $Q(71) = 152.76$, $p < .001$, $I^2 = 57.19\%$.

Figure 6

Forest Plot showing the Outcomes of 27 Individual Studies (n = 72 effect sizes) examining the Zero-Order Relation between SOFB and EF in Children



Note. WM = Working memory, IC = Inhibitory control.

To identify potential sources of heterogeneity among the effect sizes of the child studies, we conducted a series of exploratory analyses on potential moderators. The results of these analyses are presented in Table 1. The first moderator considered was children's culture. Given insufficient number of studies from each of the 16 individual countries, we aggregated studies by Western versus Eastern cultures. As shown in Table 1, SOFB and EF were significantly correlated within each culture, and did not vary significantly between cultures $F(1,23) = .63$, $p = .436$. We then investigated whether specific executive domains (i.e., attention, inhibitory control, planning, cognitive flexibility, working memory) moderated the overall effect size. Given the low number of effect sizes reported for planning ($k = 2$), this executive domain was omitted from the moderator analysis (Weisz et al., 2017). As shown in Table 1, all remaining domains were significantly and similarly associated with SOFB. Consistent with these similar effect sizes, the overall effect of executive domain, did not significantly moderate the outcome, $F(3,23) = .67$, $p = .577$.

Task-related variables were assessed next. We analyzed the effect of 1) the specific nature of the EF/SOFB measures and 2) the presentation order of these measures. As mentioned at the outset, four EF tasks were most frequently used across studies: Digit span, Word span, Stroop Colour/Word and DCCS. As shown in Table 1, these all revealed near equivalent associations with SOFB, and did not differ significantly, $F(3,17) = 2.49$, $p = .072$. Similarly, no significant differences were found across the type of SOFB task used, $F(2,24) = .66$, $p = .527$. All of the task types were significantly related to EF and the differences across tasks were not large. Finally, whether a study administered the different EF and SOFB task batteries in fixed, randomized, or mixed order did not have a significant influence on the outcome $F(3,23) = .52$, $p = .669$.

Table 1

Categorical Moderator analyses for Correlations between SOFB and EF in Child Samples

		Mean effect			Heterogeneity Statistics			
Categorical variable	*K*	*r*[95% CI]	*SE*	*p*	I^2(%)	*QE*	*df(Q)*	*QEp*
Culture	72			.436	56.90	146.27	70	< .001
Western	60	.24[.17, .31]	.03	< .001				
Eastern	12	.24[.12, .52]	.09	< .001				
EF Domain	70			< .001	59.06	148.99	66	< .001
Attention	5	.29[.19, .38]	.05	< .001				
Cognitive flexibility	14	.25[.17, .34]	.04	< .001				
Inhibitory control	20	.23[.13, .32]	.05	< .001				
Working memory	31	.27[.19, .36]	.04	< .001				
EF measure	41			< .001	57.46	143.35	67	< .001
Digit span	19	.29[.19, .39]	.05	< .001				
DCCS	6	.29[.23, .37]	.02	< .001				
Stroop	10	.28[.19, .37]	.04	< .001				
Word span	6	.28[.13, .43]	.07	< .001				
SOFB measure	72			< .001	57.88	149.99	69	< .001
Perner & Wimmer	46	.20[.19, .37]	.05	< .001				
Sullivan et al.	20	.22[.10, .34]	.06	.001				
Astington et al.	6	.22[.14, .29]	.04	< .001				
Task order	58			.669	58.28	146.10	68	< .001
Fixed	21	.22[.16, .27]	.03	< .001				
Randomized	29	.23[.10, .36]	.06	.001				
Mixed	8	.35[.12, .57]	.11	.064				

Note. *k*= number of effect sizes, SE = Standard error of the mean, *r* = Pearson correlation between SOFB and EF, 95% CI = 95% confidence intervals of the effect size, *Q*E = statistic of residual heterogeneity, *Q*E*p* = *p*-value of residual heterogeneity.

In addition, we assessed the potential moderating effects of five continuous variables. As seen in Table 2, these variables included age, sample size, presence of SOFB control questions, and numbers of SOFB trials, SOFB measures, and EF measures administered. Each of these variables was entered as a covariate in five independent meta-regressions. Two of these revealed significant effects - number of SOFB measures and number of SOFB trials. First, with respect to the size of the SOFB test battery, studies administered either one or two tasks. Our analysis revealed that including an additional measure in a study yielded significantly larger effect sizes $F(1,25) = 8.79$, $p < .007$. Second, even in cases where a single type of SOFB task was administered, some studies presented multiple trials of the same task (with some slight variations in the scenario or characters involved, for example). Including more SOFB trials had a significant impact on the overall outcome $F(1, 25) = 11.70$, $p < .002$.

Table 2

Continuous Moderator Analyses of the Correlation between SOFB and EF in Child Samples

	Mean effect						Heterogeneity Statistics		
Continuous Variables	*B*	*β*	SE	*t*	95% CI	*p*	I^2(%)	*QE*	*QEp*
Age	.008	.193	.02	.47	[-.028, .044]	.642	56.79	147.06	< .001
Sample size	-.001	.316	.00	-1.87	[-.001, .001]	.073	54.02	134.88	< .001
SOFB control questions	-.110	.411	.05	-2.03	[-.222, .002]	.053	54.46	148.47	< .001
Number of SOFB trials	.116	.042	.04	3.02	[.049, .197]	.002	44.88	124.05	< .001
Number of SOFB tasks	.155	.019	.05	2.97	[.047, .262]	.007	44.65	121.65	< .001
Number of EF tasks	.025	.187	.02	1.73	[-.005, .056]	.096	56.02	148.93	< .001

Note. *B* = change in the effect size for every one-unit change in the moderator, SE = Standard error of the mean, CI = 95% confidence intervals of the effect size, *Q*E = statistic of residual heterogeneity, *Q*Ep = *p*-value of residual heterogeneity, *d*f(*Q*) = 70 for all variables.

We also conducted a multilevel meta-regression seeking to further examine the relative contribution of these two significant moderators on the relation between EF and SOFB. This regression model was significant $F(2,25) = 44.86$, $p < .001$, but only the number of SOFB trials was a significant predictor of effect sizes, $r = .08$, CI [.01, .16], $p < .045$. Including these moderating variables into the model had the effect of reducing heterogeneity in effect sizes from 57.19% to 39.51%. However, the residual heterogeneity remained significant, $Q(69) = 112.29$, $p < .001$.

Partial Correlations

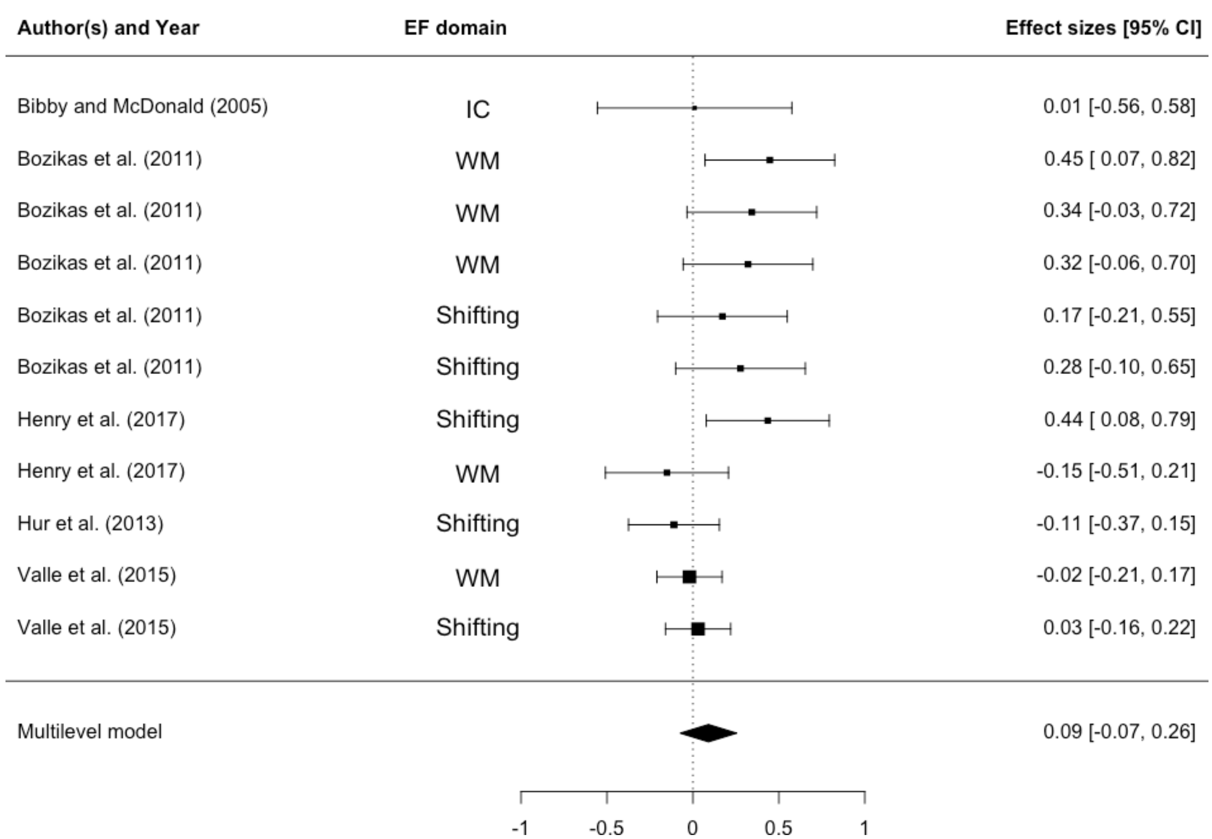
Thirteen of our meta-analyzed studies reported partial correlations accounting for extraneous factors such as participants' age or some aspect of IQ. These factors varied, however, across our dataset, and studies differed in the metrics provided. Some studies controlled for participants' age and verbal IQ for instance, while others reported partial correlations controlling for verbal IQ, non-verbal IQ, or full-scale IQ only. Nonetheless, five studies ($k = 7$) included partial correlations controlling for age only, allowing us to perform a subsample analysis accounting for this variable. Doing so also made possible the inclusion of a study that was initially not retained for analysis (i.e., Kretschmer et al., 2014) because zero-order correlations were not available. This additional analysis revealed that the association between SOFB and EF held up when children's age was held constant $r = .17$, CI [.13, .21], $p < .001$, with very low heterogeneity across effect sizes, $Q(6) = 3.84$, $p = .699$, $I^2 = 2.29\%$. This finding controlling for age within study is consistent with the finding that age across studies was also non-significant (Table 2).

Adult Samples

Figure 7 displays the forest plot of weighted mean effect sizes and confidence intervals from studies looking at zero-order associations between SOFB and EF in adults. Effect sizes ranged between r = -.15 and r = .45, with only two effects sizes significantly different from zero. Correspondingly, the overall effect size was small and not significant r = .09, CI [-.07, .26], p = .336, and showed moderate but non-significant heterogeneity in effect sizes, $Q(10) = 17.01$, p = .074, $I^2 = 44.89\%$.

Figure 7

Forest Plot showing the Outcomes of Five Individual Studies (n= 11 effect sizes) examining the Zero-Order Relation between SOFB and EF in Adults



Note. WM = Working memory, IC = Inhibitory control.

In keeping with our child samples, we conducted a planned moderator analysis. The limited number of available effect sizes for this age group, however, restricted our analyses to a single categorical variable: executive domain (Table 3). Given that no study examined how SOFB skills might be related to attention or planning abilities, and that only one effect size was reported for the relation between SOFB and inhibitory control, this analysis compared the effect sizes of two executive domains only (i.e., cognitive flexibility and working memory). As shown in Table 3, neither was significantly associated with SOFB.

Table 3

Categorical Moderator analyses for the Correlation between SOFB and EF in Adult Samples

		Mean effect			Heterogeneity statistics			
Categorical variable	*K*	*r* [95% CI]	SE	*p*	I^2(%)	*QE*	d*f*(*Q*)	*QEp*
Executive domain	10			.632	45.31	13.34	8	< .001
Cognitive flexibility	5	.10[-.43, .64]	.12	.377				
Working memory	5	.05[-.56, .65]	.11	.715				

Note. *k*= number of effect sizes, SE = Standard error of the mean, *r* = Pearson correlation between SOFB and EF, 95% CI = 95% confidence intervals of the effect size, *Q*E = statistic of residual heterogeneity, *Q*E*p* = *p*-value of residual heterogeneity.

Finally, as with the child data, several continuous moderators were also tested: sample size, number of SOFB trials, and number of EF tasks included in a given study. As reported in Table 4, differences in sample sizes did not moderate the strength of the relation between EF and

SOFB. However, as with the child data, the effect sizes were significantly larger when participants' scores were derived from a larger number of SOFB trials, $F(1,2) = 42.34$, $p < .023$. Similarly, effect sizes also varied significantly depending on the size of the EF test battery. Including more tests produced larger effect sizes, $F(1,3) = 58.49$, $p < .005$. The final meta-regression model, however, including the number of SOFB trials and the number of EF tasks as covariates, was not significant, $F(2,1) = 42.67$, $p = .108$, $Q(7) = 6.89$, $p = .439$, $I^2 = 1.51\%$.

Table 4

Continuous Moderator Analyses for the Correlation between SOFB and EF in Adult Samples

	Mean effect						Heterogeneity Statistics		
Continuous Variables	*B*	*β*	SE	*t*	95% CI	*p*	I^2(%)	*QE*	*QEp*
Sample size	-.002	.232	.01	-1.68	[-.01, .00]	.192	38.98	11.99	.214
Number of SOFB trials	.094	-.082	.01	6.51	[.03, .16]	.023	2.87	7.49	.485
Number of EF tasks	.098	-.181	.01	7.65	[.06, .14]	.005	4.73	8.35	.499

Note. *B* = change in the effect size for every one-unit change in the moderator, SE = Standard error of the mean, 95% CI = 95% confidence intervals of the effect size; Q*E* = statistic of residual heterogeneity. Q*Ep* = *p*-value of residual heterogeneity, *df*(*Q*) = 8 for all variables.

Discussion

The primary goal of this meta-analysis was to provide a summary estimate of the shared variance between SOFB and EF across typically-developed individuals from a wide age range. We employed multilevel modelling techniques to estimate the pooled effect size of over 80 correlation coefficients, extracted from both school-age children and adult samples. Considering the associations with EF consistently found in the first-order ToM domain, combined with evidence of long-lasting effects of EF in later-developing and more advanced ToM, we expected our analyses to show an overall significant association between individual differences in SOFB understanding and EF.

The overall estimate concurred with this prediction; studies of adults and children jointly revealed a small-to-moderate association between SOFB and EF (r = .23). A noteworthy consideration, however, is that the degree of overall variance explained by EF was driven in large part by studies of *children's* SOFB. A caveat, however, is that adult and child data differed substantially in their degrees of interpretability; adults only represented a small proportion of the meta-analyzed data, and many moderation effects could not be assessed systematically in this subset. For these reasons we provide individual interpretations of each age group's results before discussing the overall significance of our findings.

Relation between EF and SOFB in Children

For children there was clear evidence of positive association between SOFB understanding and EF (r = .25). The strength of this association was not only stable over a period spanning the early (5.5 years) to middle childhood years (12 years), but it also resisted statistical adjustments for age (r = .17). Further, our moderator analyses suggested some degree of cultural universality in the relations between SOFB and EF. The association between EF and SOFB did not differ on the basis of children's cultural background (i.e., Eastern or Western) across 16 countries represented in this meta-analysis.

Heterogeneity based on study characteristics reflected two important psychometric contributors to the relation between SOFB and EF. These contributors concerned the variance in the SOFB test battery administered to children. Studies in general included a limited number of SOFB assessments (i.e., one or two individual tests), although our analysis showed that even a small increase in the number of tests had a positive effect on the outcome. Relative to single-test designs, expected to yield limited scoring variability, larger effect sizes emerged from studies that assessed SOFB across a larger number of tests. Similarly, some studies presented multiple trials of the same SOFB task, which also presumably contributed to the likelihood of detecting inter-individual variability, and resulted in larger effect sizes.

A recurrent theme in the study of EF in relation to ToM is the relative variance shared by different executive components and the efficiency with which one reads others' mental states. With respect to the current meta-analysis, the available data allowed for the comparison of four EFs: attention, working memory, inhibitory control and cognitive flexibility. Each of these EFs contributed significantly, although relatively equally to SOFB performance. One explanation for the lack of specificity found here is that, in childhood at least, various EF components may make comparable contributions to SOFB. SOFB as a skill may require input from more diverse, less specialized EFs (Hughes & Devine, 2015). As mentioned earlier, it is increasingly believed that preschoolers' inhibitory abilities independently predict FOFB comprehension, especially when the task concurrently taps working memory (e.g., Carlson & Moses, 2001). The inferential processes involved in SOFB appeared to relate to EF with less specificity in the meta-analysis. However, individual studies placed different emphasis on the EF constructs of interest - some test batteries focused on a single EF, and few studies evaluated EF across its multiple constituents. Even pooled, the number of effect sizes analyzed under each EF remained relatively low, and although the size of the EF battery did not appear to influence the strength of the SOFB-EF association across studies, additional and more comprehensive research is needed to obtain clearer conclusions regarding the specificity of this association.

Relation between EF and SOFB in Adults

The association of EF to SOFB was much less evident in this age group. Overall, EF was related to SOFB only to a small, non-significant extent ($r = .09$). However, before concluding that EF and ToM are unrelated in adults, a number of caveats need to be considered.

First, with only one exception (i.e., Valle et al., 2015), our effect sizes were derived from control group data included as part of studies examining the broader concept of social cognition in adults with schizophrenia, brain injury, or multiple sclerosis, for example. Whether these data would generalize beyond their intended purpose is significantly challenged by the fact that 1) control group selection is often based on specific matching criteria (e.g., age, gender, medical history, education), and 2) some task-related procedures designed specifically for clinical populations, such as adapted language or reduced complexity, may have affected performance variability in the control samples (Chan et al., 2008; Cooper et al., 2017; Snyder et al., 2015).

Second, and relatedly, the absence of available age-adapted SOFB measures meant that the tasks presented to adults were initially devised for 5-10-year-old children (e.g., Perner & Wimmer, 1985). These tasks, in their original format, may suffer from limitations when administered to older samples - e.g., they are conveyed via child-oriented language, scored dichotomously, and not always suitable for multiple-trial assessments. Some important variability may have been masked by these limiting factors, and stronger effects may emerge given more sensitive, age-appropriate testing.

Third, given the small number of effect sizes, the available data were insufficient to conduct a fine-grained analysis of potential moderating factors. The association of SOFB with inhibitory control, for instance, was only reported once in the adult literature. Bozikas et al. (2011) found that individual scores on a Stroop Colour/Word task moderately correlated with SOFB performance (r = .45). Unfortunately, the moderating influence of this EF domain could not be considered in our analysis, which had to be restricted to just two domains (i.e., working memory and cognitive flexibility). Bozikas et al.'s study was also the only one administering more than two EF tests, which could explain that in contrast with child studies, the size of the EF battery did moderate the outcome in this age group. Finally, the effect of additional moderating variables based on sample or study features remain largely unknown given the current gaps in the adult literature linking SOFB and EF.

Implications of Meta-Analytic Findings across Age Groups

It is noteworthy that the overall effect size observed across both age groups remained relatively small. Notably, it was smaller than the one reported in the Devine and Hughes (2014) meta-analysis (r =. 44), which generally sampled younger children. Our samples did not overlap with Devine and Hughes, and while the smaller effect size observed here could be the result of different methodologies across studies, or the fact that some of Devine and Hughes' effect sizes were calculated on the basis of composite FOFB/SOFB scores, it is also possible that older children rely on EF for mental state attributions to a lesser extent than younger children. A few studies of first-order ToM in middle childhood, for instance, have found that although still significant, the association with EF tends to weaken as children progress through the early school years (e.g., Austin et al., 2014; Devine et al., 2016). There is, notably, maturation of the prefrontal cortex occurring during this period (Luna et al., 2001). With these changes, children may be able to exert more control over their thoughts and actions, and rely less on EF for belief attributions. Additionally, many authors explain these weaker correlations by referring to the type of everyday experiences provided by school settings, combined with other social factors such as friendships (Banerjee et al., 2011), and conversations about the mind (Lecce et al., 2014), which have a significant influence on individual differences in ToM during this developmental period (Hughes & Cutting, 1999; Hughes et al., 2005; Ronald et al., 2006). Similar factors could also offer a ready explanation for the weak association found in adults, whose rich social lives may contribute to lessen the importance of EF to ToM (Di Tella et al., 2020).

However, in line with some of the limiting factors discussed earlier, the overall weak strength of association observed here could also be explained by issues of sensitivity present across child and adult studies. Ceiling performance on SOFB tasks, for instance, was suspected in a number of cases, but could not be assessed directly due to a lack of available information (i.e., performance means and/or range were not included in individual articles). Low variability in task scores could in some instances be presumed, however, from the type of dependent measures used to index EF and/or ToM (e.g., narrow nominal scales), as well as the very limited range of age-adapted measures present across adult studies in particular. It is known that in children and adults over the age of approximately 6 years, individual differences are often best captured on the basis of response times, or processing speed, rather than response accuracy only (e.g., Cooper et al., 2017; Davidson et al., 2006; Sternberg, 2010). Yet, of the 32 studies included for analysis, only two used participants' response times as an index of EF performance, and none of the studies used this outcome measure to index SOFB performance. In general terms, measurement sensitivity is a crucial component of individual differences research (Sternberg, 2010), as supported here by the fact that the number of trials as well as the size of the test battery were significant moderators of the relations between SOFB and EF. Increasing measurement sensitivity may be a potential way forward for studies examining how ToM is linked to EF in older children and adults.

Collectively, these factors limit the conclusions that can be drawn from the effect sizes reviewed in the present meta-analysis. Although the meta-analysis of the studies conducted in school-age children confirmed significant links between SOFB and EF, less confidence can be placed in the findings for adults. In Study 2, we follow up on these limitations, by introducing new age-appropriate methodologies. In particular, we devised a novel SOFB paradigm, and adopted a much more extensive individual differences approach than previously used to investigate the links between SOFB and EF in normally-developed adults.

CHAPTER 5

Study 2

Individual Differences in Second-Order False-Belief Reasoning and Executive Processes in Adults: An Investigation into the Nature and Magnitude of Associations

As uncovered in Study 1, the scope of the extant research linking second-order ToM and EF in adults is quite narrow and of limited interpretability. In contrast to children, the reviewed findings pointed to substantial gaps in our knowledge of whether and to what extent second-order ToM relates to EF in adults. While there remains the possibility that for adults, performance on SOFB tasks is, indeed, independent of EF, this would be largely inconsistent with evidence of ongoing reliance on EF for mental state reasoning across the lifespan (e.g., Bradford et al., 2015; German & Hehman, 2006; Lin et al., 2010; Wardlow, 2013).

Instead, it may be that if tested under more appropriate conditions, significant associations may be detected in adulthood as well. To this end, the present study examines the degree to which second-order ToM covaries with EF in typically-developed adults, given more sensitive, age-appropriate methodologies.

Previous Evidence: Summary of the Main Limitations and Research Gaps

In Study 1 many factors limited our understanding of the relation between second-order ToM and EF in adults. First, only one of the five meta-analyzed studies provided data for the joint assessment of SOFB and EF from a standard sample of neurotypical adults (i.e., Valle et al., 2015). All other effect sizes were obtained from control group data included in the context of clinical research with the primary goal of assessing, not the relations between SOFB and EF, but the more general cognitive functioning of a given patient group. Second, most studies administered a single EF or SOFB assessment, which may have caused low variability in task scores, thereby limiting the power to detect individual differences in performance (Chan et al., 2008; Cooper et al., 2017; Snyder et al., 2015). Third, and relatedly, the majority of adults were tested with paradigms initially developed for school-age children, presented in their original child-directed language, and scored according to their original, narrow scales (i.e., scores typically ranging between 0-2 points). Without modifications, these tasks are likely to produce ceiling effects in older samples, and hence a large degree of uniformity in performance (Birch & Bloom, 2007; Bloom & German, 2000).

Fourth, beyond the fact that very few individual studies were available for analysis, the specificity of the SOFB-EF relation remains undetermined. The association of SOFB with inhibitory control, for instance, was only reported once in the adult literature (Bozikas et al., 2011), and no prior research has jointly examined the effect of inhibitory control, working memory, and cognitive flexibility within the same study. Finally, there was very limited control for/or investigation into potential sources of extraneous variability in the relation between SOFB and EF. Even pooled across studies, the available data were insufficient for a reliable analysis of the moderating effects of study quality factors - e.g., control questions, order of task administration - or other more general aspects of intelligence or cognitive functioning.

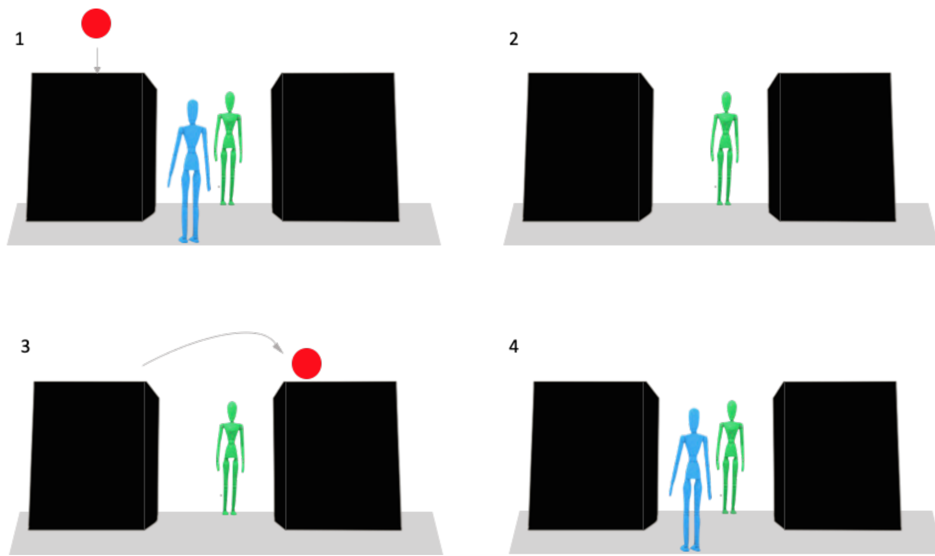
The Current Work

The current study was designed to address the limitations of existing research through an in-depth assessment of the relations between SOFB and EF. First, we designed a new age-appropriate alternative to the classic SOFB paradigm, adapted from earlier work by Apperly et al. (2006a; Apperly et al., 2006b; Back & Apperly, 2010). In contrast to more traditional approaches in which SOFBs are induced by adding a number of story elements to the first-order task, we created a task with a recursive thought structure that kept the task demands as equivalent as possible for first- vs second-order processing and for true- vs false-beliefs.

The experimental stimuli for the FOFB task consist of a series of animated clips depicting two opaque containers (left or right), an object (red ball) and two agents (“Blue” and “Green”). Participants observe as the object is placed in either one of the containers, in the presence of both agents. One of the agents (e.g., “Blue”) then exits the scene and, unbeknownst to her, the object transfers from its original location to the alternate container. Upon her return to the scene, the participant is asked to select the container in which the agent believes the object to be housed (see Figure 8).

Figure 8

A Schematic Depiction of the Sequence of Events for the FOFB Task

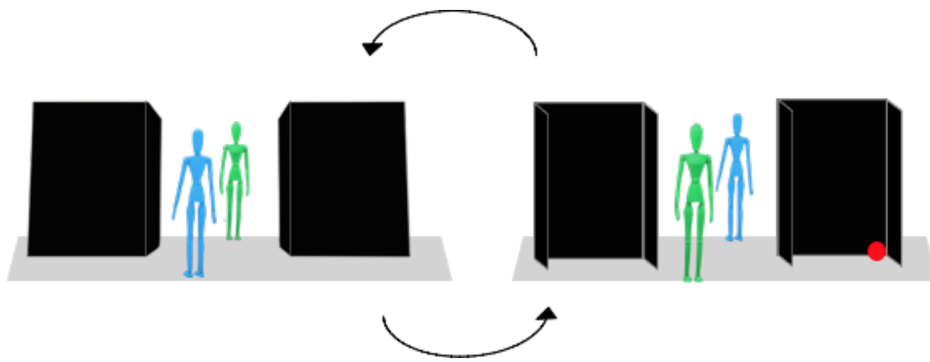


Note. Both agents are watching as a red ball enters one of the boxes (1), Blue agent leaves the room (2), the ball moves from its original location to the alternate location in the absence of the Blue agent (3), Blue agent returns (4).

The SOFB task follows the same logic as the FOFB task, but with two further modifications made to Apperly et al.'s (2006a) original paradigm. Primarily, we introduced containers with a transparent back panel. This resulted in a perspective mismatch between the agents in the scene: The containers were entirely opaque from Blue's standpoint, but were see-through from Green's standpoint (see Figure 9).

Figure 9

Still Frames from an Introductory Video Demonstrating the Transparent Feature of the Boxes

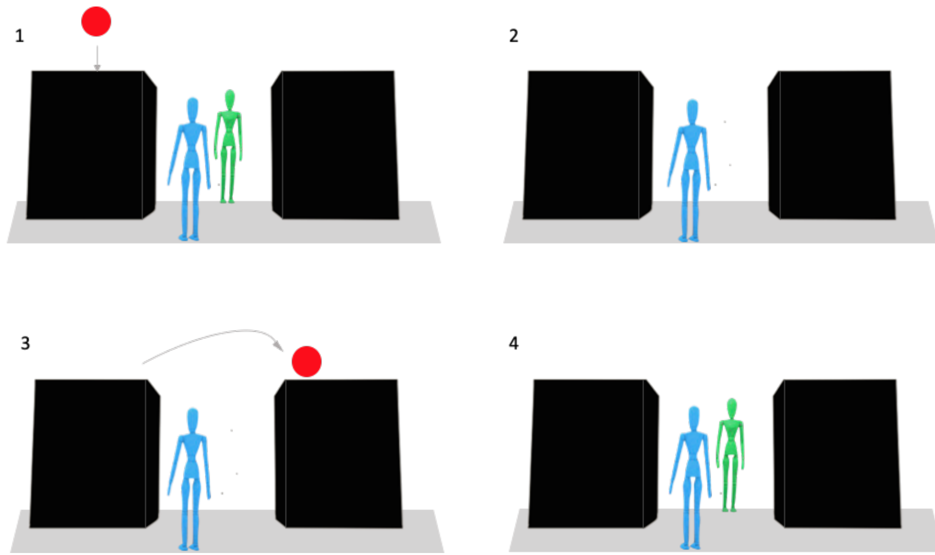


*Note***.** The panel on the left shows that the boxes are entirely opaque from the Blue agent's (and the participant's) perspective and the panel on the right (rotated 180 degrees) shows that the boxes are see-through from Green agent's perspective.

With this small but crucial change, Green had permanent visual access to the content of the containers, such that even in situations where the object switches locations in her absence (i.e., false-belief events), her belief states are always updated upon her return in the scene. Importantly, participants are also told that Blue cannot see over the containers, and thus does not know that the back is see-through. This manipulation has the effect of inducing a SOFB, where Blue does not know that Green knows where the ball is, and thus holds false-beliefs about Green's beliefs (see Figure 10).

Figure 10

A Schematic Depiction of the Sequence of Events for the SOFB Task



Note. Both agents are in the room while a red ball enters one of the boxes (1), Green agent leaves the room (2), the ball moves from its original location to the alternate location in the absence of the Green agent (3), Green agent returns (4).

Hence, we extracted the key logical features of the Ice Cream task, but stripped down its components so that, *aside from the representational content they assess*, the information processing requirements that the first and second-order tasks impose are very similar. Recall, for instance, the elements of Wimmer and Perner's (1985) Ice Cream Truck task: There are four agents (John, Mary, Ice Cream Man, Mary's mother), three locations (park, church, Mary's house), as well as several story elements (the story unfolds over a long narrative sequence of events and contains many different parts). In our modified task, the number of possible locations was reduced from three to two, as in the first-order case, and so was the number of belief-induced agents. Additionally, the task was carefully sequenced to be limited to only two main events, and designed to precisely match the duration of the first-order task (14000 milliseconds).

Second, to obtain a comprehensive understanding of the variance that second-order ToM may share with EF, participants' performance on this new task was evaluated against their performance on multiple tasks representative of the three main components of EF (i.e., inhibitory control, working memory, and cognitive flexibility). A certain risk of "task impurity" is inevitable in EF research, given that specific tasks are likely to trigger more than one EF (Miyake et al., 2000b; Zelazo et al., 2016). Including multiple tests, however, is recognized as good practice to minimize this risk (Best & Miller, 2010; Hughes & Graham, 2002; Miyake et al., 2000a), and also contributes to increasing measurement precision while maintaining variability in test scores (Crane et al., 2008; Snyder et al., 2015).

Third, a further step to ensure measurement sensitivity was the use of outcome measures able to detect subtle differences in performance. Where possible, we used measures of participants' reaction times as the main dependent variables across both ToM and EF tasks. Reaction times provide sensitive indicators of performance, and have the ability to prevent the occurrence of floor/ceiling effects (Draheim et al., 2016; Holden et al., 2019; Magnus et al., 2019).

Finally, given that individual differences in ToM are associated with individual differences in other domain general skills conceptually related to ToM, we included two measures of verbal ability as control variables. Language has been shown to be an important factor contributing to the development of both first-order ToM (Milligan et al., 2007) and second-order ToM (Arslan et al., 2017; Flobbe et al., 2008; Hollebrandse et al., 2014;

Polyanskaya et al., 2019), yet no study to date has examined its relation to EF and ToM in adulthood.

Study Aims and Predictions

The present study thus examined whether the EF-SOFB relation observed in middle childhood persists through adulthood, given methods of increased sensitivity and age-appropriateness. We sought to address two issues:

(1) If EF continues to play a role in the SOFB skills of fully-developed individuals, then participants with higher performance in the EF tasks should show more efficient SOFB processing than participants with lower EF levels. Alternatively, if the role of EF is restricted to simpler FB concepts, or to the SOFB skills of younger individuals, our analyses should not reveal significant SOFB-EF associations.

(2) Our earlier task analysis of SOFB provided reasons to think that the three main components of EF could be implicated in second-order ToM. However, we do not have a direct hypothesis pointing to which specific executive components should share the most variance with SOFB. There may be continuity in the role of EF across FOFB and SOFB, such that the particular EFs associated with FOFB would also make unique contributions to SOFB (Leslie et al., 2005; Perner et al., 2002). In the event that discontinuity is observed, such that some EFs are important for FOFB, but not for SOFB understanding, and vice-versa, then this would suggest an executive distinction between FOFB and SOFB (Apperly et al., 2009; Miller et al., 2018).

Our research plan and hypotheses for this study were pre-registered on the Open Science Framework at osf.io/vf7dn.

Methods

Participants

One hundred undergraduate students (68 female) from Victoria University of Wellington, New Zealand, took part. Participants had an average age of 20 years (Range: 17-63). All were fluent in English with normal or corrected-to-normal vision. In this and the following study, signed consent was obtained from all participants and ethical approval was granted prior to commencement.

Measures

Measures included assessments of first and second order true and false-belief reasoning, EF, as well as verbal ability as a control. The choice of measures was informed by previous research with adults and selected to facilitate comparison with previous studies. Specifically, each measure was chosen based on criteria of sensitivity to variation in performance and acceptable age-related psychometric properties (e.g., internal consistency $\alpha > .70$, test-retest reliability $r > .70$). All EF tasks were computerized versions of standard neuropsychological tests implemented by the open-source Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014) and, as is conventional in individual differences research, were administered in a fixed order as follows, along with false-belief and verbal ability measures: Letter fluency, Lexical decision, FOFB, SOFB, Trail Making Test, Plus/Minus, Stroop Colour/Word, Simon task, Backward Digit Span, and *N*-Back.

Verbal Ability Measures

As measures of verbal ability, we included two indices of language-based reasoning: a letter fluency task (Borowski et al., 1967) and a lexical decision task (Meyer & Schvaneveldt, 1971). Letter fluency assesses lexical retrieval ability and consists in generating as many words as possible beginning with a target letter "C", and then a target letter "S". Scores are calculated on the basis of the total number of unique words identified under a time constraint of 60 seconds per letter. Lexical decision is a word recognition test in which participants must decide whether a string of letters represents an English word (e.g., "judge") or a nonword (e.g., "ludge"). There are 58 trials, and decision time is used as the dependent variable.

False-Belief Measures

First-order false-belief: As described above, this task is modelled after a modified version of the well-known "unexpected transfer" false-belief scenario for use with adult samples (Apperly et al., 2006a; Wimmer & Perner, 1983). On each trial, a belief inducing video is played, followed by a question mark, and the presentation of a test probe (e.g., "Where does Blue think that the ball is?"). The test probe is presented until a response is selected. Two true-belief (i.e., the agent witnessed the object transfer) and four false-belief events were presented in randomized order, and the correct answer was equally likely to refer to the container on the left, or the container on the right. Reaction time was recorded on every trial, from the onset of the test probe.

Second-order false-belief: This task's stimuli are identical to that of the FOFB task described above, with the exception that participants are asked to make judgments about what one of the agents thinks that the other agent believes about the location of the object (e.g., "Where does Blue think that Green thinks that the ball is?"). As a familiarization, participants commenced this task by watching a 360-degree rotation of the experimental scene showing Blue's occluded viewpoint, and Green's open viewpoint. In keeping with the FOFB task, two true-belief events and four false-belief events were presented, totaling six SOFB trials. The dependent variable of interest was participants' response speed, as in the FOFB task.

Executive Function Assessment

Cognitive flexibility measures

Trail Making Test: In part A (TMT-A), participants are instructed to connect 25 randomly positioned numbered circles in ascending order as quickly as possible. In part B (TMT-B), the task is to connect randomly positioned numbers and letters, alternating between numbers in ascending order and letters in alphabetical order (e.g., 1-A, 2-B, 3-C). Scoring is based on subtracting time to completion of the TMT-B from time to complete the TMT-A. A lower score therefore indicates better performance than a higher score (Partington & Leiter, 1949).

Plus/Minus task: In this task adapted from Miyake et al. (2000b), three different series of two-digit numbers are presented with different accompanying instructions. The first series involves adding 3 to each number presented on the screen (25 trials), while the second requires subtracting 3 from each number (25 trials). Participants are then instructed to sequentially alternate between adding 3 or subtracting 3 from the numbers on the third series (25 trials). The primary variable of interest is the shifting cost, which is obtained by subtracting the total time required to complete the addition and subtraction series from the total time required to complete the alternating series. Low shifting costs therefore reflect better cognitive flexibility.

Inhibitory control measures

Stroop Colour/Word: The first part consists of colour-naming trials where participants are presented with colour words and asked to indicate the colour of the word by key press. The second part consists of two types of trials, congruent trials where the word and the font colour are matching (i.e., the word "red" written in red font); and incongruent trials where the word and font colour are not matching (i.e., the word "red" written in blue font). There are 26 trials of each condition, for a total of 72 trials. The outcome of interest is the "Stroop effect", a measure of interference obtained by deducting the mean correct response times on the congruent trials from the mean correct response times on the incongruent trials (Stroop, 1935). Greater inhibitory efficiency is indexed by smaller Stroop effects.

Simon task: The task consists in judging the colour of a circle stimulus while ignoring its horizontal position on the screen. Participants are instructed to respond to a "blue" circle by pressing the left key, and to a "red" circle by pressing the right key. The side of the screen on which stimuli are presented influences participants' responding by either matching (i.e., congruent trials) or not matching (i.e., incongruent trials) the side of the correct key press associated with the colour of the shape. The task comprises 48 trials of each condition. The primary inhibition efficiency index is the "Simon effect", calculated as follows: mean correct response times on incongruent trials minus mean correct response times on congruent trials (Simon & Wolf, 1963). Low Simon scores are interpreted as resulting from less interference from incompatible trials, therefore indicating greater inhibitory control.

Working memory measures

Backward Digit Span: Participants are presented with sets of increasingly longer series of digits (e.g., "3-2-1-4") and asked to recall the series in reverse order (e.g., "4-1-2-3"). Presentation begins with three digits in a series, and the number of digits increases by one as participants succeed at two consecutive trials for a maximum of ten digits per series. Scores are based on the longest series of digits correctly recalled backwards (Hebb, 1961).

N-Back: The *N*-back task involves serial presentation of a sequence of numbers, and participants must judge whether each new number presented matches the one displayed *n* trials ago (up to 3-back). For instance, in the two-back and three-back conditions, the accurate response corresponds to the number presented two and three items back, respectively. Scores are obtained by calculating the proportion of correct responses across the 66 trials comprising this task (Braver et al., 1997).

Results

All statistical analyses were performed in *R* (R core team, 2019) and conducted in accordance with our preregistered protocol. As a first step, all distributions were examined to ensure that the assumptions of the analysis were met. Skewness and kurtosis of four RT-based measures (SOFB, FOFB, TMT, Lexical Decision) were improved with Yeo-Johnson transformations to approximate normality (Yeo & Johnson, 2000). All measures achieved a satisfactory level of normality after this transformation process (i.e., skewness < 2 and kurtosis < 4). In a second step, we screened the data from all tasks for the presence of outliers. Specifically, and in line with previous EF studies (e.g., Archibald et al., 2015; Rapport et al., 2002), scores for any task beyond 1.5 times the maximum and minimum values of the interquartile range were winsorized to the nearest non-outlying value. This procedure affected no more than 1.98% of the overall data. Finally, in order to assist comparison across tasks scored on different units, all dependent measures were standardized ($M = 0$, $SD = 1$). However, for ease of interpretation the numbers below and presented in figures were backtransformed to raw performance data. We also reversed the directionality of three accuracy-based variables (*N*-Back, Digit Span, and Word Fluency) such that lower values indicate better performance across all analyses. Across all false-belief and EF tasks, we analyzed reaction time data for correct responses only.

Descriptive statistics for all measures are presented in Table 5. As can be seen, all variables showed adequate variability in performance.

Table 5

Descriptive Statistics for all Measures

	Measure	DV	Mean	*SD*	Min	Max	Skew	Kurt
ToM	FOFB	RT	849.66	523.22	188.50	2523.00	-.03	-.43
	FOTB	RT	1111.24	1176.40	149.50	6486.00	.03	-.12
	SOFB	RT	967.86	767.79	200.25	3474.00	.02	-.22
	SOTB	RT	777.49	580.50	177.50	2457.50	.15	-.08
VA	Lexical Decision	RT	1203.52	313.83	707.79	2235.31	-.07	-.33
	Word Fluency	Number of words	28.81	8.67	7.00	51.00	.08	-.08
IC	Stroop	Stroop effect	119.53	118.39	-172.04	521.32	.60	.13
	Simon	Simon effect	50.63	61.68	-146.10	360.05	-.18	-.38
CF	Plus/Minus	Shifting effect	418.37	405.68	-632.86	1785.18	.48	.12
	TMT	Shifting effect	1092.87	348.75	526.89	1828.31	.28	-.72
WM	*N*-Back	Number of CR	56.36	5.83	43	65	-.80	-.21
	Digit Span	Span length	4.50	1.63	2	9	.31	-.48

Note. DV= Dependent Variable; ToM= Theory of Mind; FOFB= First-Order False-belief; FOTB= First-Order True-belief; SOFB= Second-Order False-Belief; SOTB= Second-Order TB; VA= Verbal Ability; IC= Inhibitory Control; CF= Cognitive Flexibility; WM= Working Memory; CR= Correct Responses; RT= reaction time in milliseconds; TMT= Trail Making Test. *N*= 100 except for FOFB and SOFB: *N* = 98 as two participants failed all four trials on these tasks.

Correlations Within and Across Executive Function and Verbal Ability Assessments

The zero-order correlations among all measures are displayed in Table 6. As our assessment battery consisted of three pairs of individual measures each believed to tap a specific EF dimension, we first conducted a construct-by-construct analysis examining the degree of shared variance among each pair. As Table 6 shows, there were interrelations between the two individual tasks comprising each EF dimension. Hence, these tasks were averaged to form three composite scores for each participant: inhibitory control (derived from the mean Simon effect and the mean Stroop effect), working memory (derived from accuracy scores on the *N*-Back task and Digit Span length), and cognitive flexibility (derived from shifting effects obtained from the TMT and the Plus/Minus tasks). Finally, given the significant correlation between Word Fluency and Lexical Decision, we similarly calculated a verbal ability composite score.

Table 6

Correlation Matrix for all Measures

	1	2	3	4	5	6	7	8	9	10	11
1. FOFB	-										
2. FOTB	$.45^{***}$	-									
3. SOFB	$.62^{***}$	$.46^{***}$	-								
4. SOTB	$.34^{***}$	$.38^{***}$	$.51^{***}$	-							
5. Lexical decision	$.27^{**}$	$.28^{**}$	$.26^{**}$	.07	-						
6. Word fluency	.14	.01	.16	.07	**.29*****	-					
7. Stroop effect	.03	.16	.08	$.26^{*}$	.02	$.21^{*}$	-				
8. Simon effect	.13	.08	.09	.08	.01	.16	**.28****	-			
9. Plus/minus effect	.17	$.21^{*}$	$.29^{**}$	$.21^{*}$	.19	.03	$.24^{*}$	$.26^{***}$	-		
10. TMT effect	$.31^{**}$	.14	$.32^{**}$	.19	.02	$.20^{*}$	.04	$.20^{*}$	**.23***	-	
11. *N*-Back	.04	.03	$.27^{**}$	.03	.14	$.22^{*}$	.15	.06	.05	$.29^{**}$	-
12. Digit Span	$.27^{**}$	.14	$.26^{**}$	.09	.03	.19	.15	$.21^{*}$	.14	$.44^{***}$	**.23***

Note. FOFB = First-Order False-Belief; FOTB = First-Order True-Belief; SOFB = Second-Order False-Belief; SOTB = Second-Order True-Belief; TMT = Trail-Making Test. Cells denoting pairwise correlations within a EF sub-domain are bolded. * $p < .05$; ** $p < .01$; *** $p < .001$.

Assessed next were intercorrelations among these EF composite scores, and whether EF was associated with verbal ability (see Table 7). Consistent with prior work (e.g., Friedman et al., 2008; Miyake et al., 2000b; Zabelina et al., 2019), these analyses suggested that some, but not all of these scores correlated. Specifically, as Table 7 shows, cognitive flexibility was significantly related to both inhibitory control and working memory (*ps* < .022). The correlation between inhibitory control and working memory was in the predicted direction, but failed to reach significance ($p = .071$). Further, verbal ability did not significantly correlate with any of the EF scores (all *p*s > .053).

In subsequent analyses, the relative contribution of EF to SOFB performance was examined with separate correlational and linear regression analyses for the constructs of inhibitory control, cognitive flexibility, and working memory. Given the recognized importance of considering both the unity and diversity of EF (Friedman et al., 2008; Garon et al., 2008; Miyake et al., 2000b), we also collapsed data across all three constructs for analysis of a composite EF score in relation to FOFB and SOFB performance.

Table 7

Zero-Order Correlations among Composite Measures of EF and Verbal Ability

Measure	1	2	3	4
1. Verbal Ability	-			
2. Inhibitory Control	.25	-		
3. Cognitive Flexibility	.09	.28*	-	
4. Working Memory	.18	.18	.27*	-

* $p < .05$

False-Belief Assessment

Our main objective was to assess potential links between EF and the particular ability to process SOFBs. As seen in Table 8, significant associations were obtained between composite EF and SOFB ($p < .001$). Decomposing this association into finer EF dimensions revealed that SOFB showed the highest correlation with cognitive flexibility ($p < .001$), closely followed by working memory ($p < .002$). Inhibitory control, however, was positively but not significantly associated with SOFB ($p = .737$). Results of similar significance but slightly smaller magnitude were found for verbal ability-controlled correlations (Table 8).

As shown in Table 8, SOFB was also significantly related to FOFB, and this was the case even with verbal ability partialled. We thus reconducted the previous analyses, this time accounting for variance in FOFB as well as verbal ability. Despite a slight reduction in the magnitude of associations between measures, comparable outcomes were obtained; both working memory and cognitive flexibility remained significantly associated with SOFB.

Table 8

Zero-Order and Partial Pearson Correlations between EF, Verbal Ability, SOFB and FOFB Reasoning

	SOFB	FOFB
EF composite	$.39^{***}/.39^{***}/.23^{*}$	$.29^{**}/.29^{**}$
Cognitive Flexibility	$.40^{***}/.40^{***}/.31^{**}$	$.30^{**}/.30^{**}$
Working Memory	$.31^{**}/.31^{**}/.29^{**}$	.11/.11
Inhibitory Control	.11/.10/.01	.05/.06
Verbal Ability	.03	.01
FOFB	$.62^{***}/.61^{***}$	-

Note. EF= Executive Functioning; FOFB= First-Order False-Belief. This table shows raw correlations/partial correlations controlling for verbal ability/ partial correlations controlling for verbal ability and FOFB performance.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Concerning FOFB reasoning, we conducted a similar set of analyses focusing on whether performance differences across individuals would show similar patterns of association with EF (Table 8). Although the EF composite was correlated with FOFB ($p < .001$), this effect appeared to be largely driven by the association between cognitive flexibility and FOFB ($p < .001$). Controlling for verbal ability did not significantly impact these results, and there were no further significant associations between FOFB and any of the other EF indices.

Hierarchical Regression Analysis

To assess whether each EF construct exerts independent predictive value, two separate hierarchical multiple regression models were compared with different sets of predictors and SOFB as the outcome variable. These models are summarized in Table 9.

Model 1 assessed how control variables alone (verbal ability and FOFB) predicted SOFB. In this model, only FOFB accounted for significant unique variance in SOFB performance (23.1%). The addition of the two EF variables that had significantly predicted SOFB in univariate analyses (Model 2) increased the proportion of shared variance to 33.6%, and, importantly, working memory and cognitive flexibility as well as FOFB contributed unique predictive value to SOFB.

Table 9

Hierarchical Regression Models Including all Variables Significantly Predicting SOFB

	Regression Coefficients				
Predictors	*B*	*SE*	*t*	ΔR^2	*F*
Model 1				.231	15.57^{***}
Verbal ability	.013	.021	.616		
FOFB	< .001	< .001	5.370^{***}		
Model 2				.336	13.27^{***}
Verbal ability	.007	.020	.329		
FOFB	< .001	< .001	4.981^{***}		
Cognitive flexibility	< .001	< .001	2.762^{**}		
Working memory	.056	.026	2.140^{*}		

Note. FOFB = First-Order False-Belief

* $p < .05$; ** $p < .01$; *** $p < .001$.

Exploratory Analysis of the Relations between True-Belief and EF

For our final analyses, we explored the relative contribution of EF to one additional aspect of belief reasoning: True-Belief understanding. As reported in Table 10, second-order true-belief (SOTB) was correlated with first-order true-belief (FOTB) ($p < .001$) and with composite EF ($p < .011$) - although the latter appeared to be specific to cognitive flexibility ($p < .001$). All partial correlations held up after controlling for verbal ability (ps < .013, .004, and .002, respectively, for FOTB, composite EF, and cognitive flexibility), while only cognitive flexibility significantly correlated with SOTB after the effect of FOTB was partialled out ($p < .004$). Near-identical patterns emerged from an analysis focusing on FOTB data (Table 10). Composite EF, driven by cognitive flexibility, significantly correlated with FOTB (ps < .041 and .023, for composite EF and cognitive flexibility), and both correlations held up after controlling for verbal ability (ps < .035 and .021). No further associations were significant.

Table 10

Exploratory Correlation Analysis between EF, SOTB and FOFB Reasoning

	SOTB	FOTB
EF composite	$.26^{*}/.24^{*}/.16$	$.21^{*}/.24^{*}$
Cognitive Flexibility	$.26^{*}/.25^{*}/.29^{**}$	$.23^{*}/.25^{*}$
Working Memory	.05/.04/.03	.06/.10
Inhibitory Control	.09/.07/.04	.01/.03
Verbal Ability	.08	.16
FOTB	$.38^{***}/.31^{**}$	-

Note. FOTB = First-Order True-Belief. This table shows raw correlations/partial correlations controlling for verbal ability/ partial correlations controlling for verbal ability and FOTB performance.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Next, a hierarchical regression for SOTB analogous to that conducted for SOFB was conducted. For Model 1, only FOTB accounted for significant unique variance in SOTB performance (Table 11). For Model 2, cognitive flexibility did not uniquely contribute to SOTB above and beyond the variance accounted for by FOTB.

Table 11

Hierarchical Regression Models Predicting SOTB Reasoning

	Regression Coefficients				
Predictors	*B*	*SE*	*t*	ΔR^2	*F*
Model 1				.105	6.677^{**}
Verbal ability	.032	.019	-1.699		
FOTB	.259	.079	3.265^{**}		
Model 2				.081	3.866^{**}
Verbal ability	.009	.019	.512		
FOTB	.236	.082	2.866^{**}		
Cognitive flexibility	< .001	< .001	.953		

Note. FOTB = First-Order True-Belief; FOFB = First-Order False-Belief.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Overall, SOTB appears to be best predicted by variance in FOTB understanding. This suggests that the predictive value of EF to second-order belief reasoning may be largely restricted to inferences in which a false-belief is involved. Nonetheless, as a conservative final analysis, we repeated the earlier regression predicting SOFB from EF, verbal ability, and FOFB, but this time including SOTB and FOTB (Table 12). All indices but verbal ability and cognitive flexibility accounted for significant variance in SOFB performance, and even in this analysis, working memory contributed to SOFB above and beyond the variance accounted for by true-belief reasoning.

Table 12

Hierarchical Regression Model Including all Variables Predicting SOFB

	Regression Coefficients				
Predictors	*B*	*SE*	*t*	ΔR^2	*F*
				.487	16.36^{***}
Verbal ability	.009	.017	.521		
FOFB	< .001	< .001	3.354^{**}		
FOTB	.190	.081	2.365^{*}		
SOTB	.413	.092	4.473^{***}		
Cognitive flexibility	< .001	< .001	1.209		
Working memory	.064	.023	2.834^{**}		

Note. FOFB = First-Order False-Belief; FOTB = First-Order True-Belief; SOTB = Second-Order True-Belief

* $p < .05$; ** $p < .01$; *** $p < .001$.

Discussion

Although adults in general are quite well-versed at making inferences about others' minds, thinking through "what she believes that he believes" is a non-trivial challenge known to induce important cognitive costs. Joined by empirical observation of associations in school-age children, influential work in its first-order counterpart provided a strong rationale for suspecting that certain executive skills may account for some of the constraints in individual performance on SOFB tasks. Here, we focused on a general mapping of this ability onto a broad EF assessment. We investigated the role of working memory, inhibitory control, and cognitive flexibility in explaining individual differences in SOFB understanding. We also developed experimental stimuli of increased sensitivity, to capture subtle variability in adults' SOFB performance.

We hypothesized two alternative possibilities. First, if there is ongoing reliance on EF for SOFB reasoning through adulthood, then we should find positive correlations between the two constructs. Alternatively, if EF is important for the development of SOFB in childhood, but less so for the maintenance of these skills later in life, then EF and SOFB should be only weakly related or unrelated in adults. Our findings favoured the former possibility. Participants most proficient at EF also performed well on measures of SOFB. Specifically, both working memory and cognitive flexibility correlated with SOFB performance, independently of variance accounted for by FOFB and verbal abilities. Inhibitory control, however, failed to correlate significantly with SOFB performance. Below, we review these findings in more depth, and further discuss how the complexity of a belief inference (i.e., FOFB/SOFB), as well as the type of belief held by an agent (i.e., true/false belief) may influence the contribution of EF to mental state representations.

Developmental findings often report an important role for working memory in the emergence of first-order ToM in the early years of life (e.g., Carlson et al., 2002), but also in school-age children's SOFB skills (Kretschmer et al., 2014; Perner et al., 2002). Our results suggest that this association may persist beyond childhood, but only when considering the latter type of belief attributions. Working memory, in fact, was found to share a low-to-moderate, but unique amount of variance with SOFB. As hierarchical regression analyses indicated, this association remained even with all other executive (i.e., cognitive flexibility and verbal ability) and ToM factors (i.e., FOFB/FOTB, SOTB) held constant. In contrast, no relation was found between FOFB and working memory. It may be that FOFB is sufficiently high in working memory demands to elicit individual differences in children, but insufficiently taxing for individuals with more mature ToM/EF. For these latter individuals, correlations would perhaps be expected to emerge only if a more demanding working memory threshold is imposed, such as in the case of SOFB.

Cognitive flexibility, on the other hand, showed significant associations with both FOFB and SOFB. While cognitive flexibility is less studied in relation to adult ToM, and although we can only speculate on the mechanisms giving rise to these associations, alternating among different representations of the same visual scene is an aspect of cognition well known to be resource-demanding (Ferguson et al., 2017; Kessler & Thomson, 2010; Qureshi et al., 2010). A number of perspective-taking studies show that in some instances, even the simple visuospatial perspectives of others are calculated with some effort, especially conflicting perspectives which must be generated with some mental flexibility. In such studies, delayed responses are observed when adults are asked, for instance, to describe a simple 3D environment as seen by another individual (e.g., number of objects on a wall), but also when subsequently asked to do the reverse operation of judging the scene from their own perspective. That is, whenever a perspective shift occurs within a trial, whether allocentric of egocentric, adults appear to rely on some cognitively costly processes to make the appropriate task-related judgment (Ferguson et al., 2017; Long et al., 2018; Qureshi et al., 2010; Ryskin et al., 2015; Samson et al., 2010; Samuel et al., 2018; Surtees & Apperly, 2012). Considering this evidence, it may not be surprising that cognitive flexibility predicted performance on both FOFB and SOFB, given that these two kinds of tasks require integrating multiple conflicting representations of a single event.

Further, cognitive flexibility was the only EF significantly predicting performance across specific belief contents; both FOTB and SOTB were moderately related to cognitive flexibility. Though much less pronounced, true-belief reasoning still makes some demands in ToM skills insofar as it requires tracking the content of an agent's mind, with the exception that the agent is not mistaken (Apperly et al., 2011; Döhnel et al., 2012; Friedman & Leslie, 2004; 2005, Leslie & Polizzi, 1998, Leslie et al., 2005; Phillips et al., 2011). It may be that mentally conceptualizing the world as represented by another is sufficiently taxing of cognitive flexibility to reveal individual differences on its own, without the need to resolve a FB conflict per se. While this is plausible, we treat this interpretation cautiously as it is uncommon to find associations between true-belief data and EF (e.g., Pesch et al., 2020; Phillips et al., 2011), and indeed these associations to EF were weaker for first-order than for second-order ToM. More evidence will be needed to obtain a clearer picture of the links between these two constructs.

Lastly, we found that inhibitory control was unrelated to SOFB. Given that participants' FOFB judgements were also unrelated to their inhibitory control abilities, it could be that the relationship between these two constructs, at least as measured here, does not extend through adulthood. Progress in inhibitory control may provide a foundation for children to learn the initial requirement of self-perspective inhibition, allowing them to focus more accurately on how beliefs, desires, and intentions affect others' behaviour in predictable but subjective ways. Selecting the relevant information from a belief-inducing story is likely to be a problem to overcome over the course of development, but may perhaps not be as predictive of false-belief performance once certain ToM concepts are more firmly acquired. Inhibitory control is still predictive of children's FOFB understanding in the early years of schooling for instance (Marcovitch et al., 2015), but not later in middle childhood (Austin et al., 2014; Lecce et al., 2017), and it appears from our findings that inhibitory control does not contribute to either first or second-order ToM in adults.

Collectively, our findings raise the possibility that what is involved in the process of identifying an agent's beliefs, versus what is involved in the process of using that information recursively to identify an agent's belief about another agent's belief, correspond to cognitive challenges of strongly related, but not entirely identical kinds. We did in fact observe that first and SOFB appeared to be closely related constructs (r = .61), although FOFB only correlated with cognitive flexibility, while SOFB correlated with both cognitive flexibility and working memory. It is unclear exactly what might underlie these differences, but we suggested here that for individuals with more mature EF and rich social cognitive skills, working memory may be especially important when negotiating ToM problems that meet a certain threshold of difficulty, such as SOFB. Yet, we do not exclude the possibility that cognitive flexibility may have an important influence on adults' SOFB performance. Finer-grained analyses of how working memory and cognitive flexibility each contribute to SOFB will be needed to further explore the mechanisms underlying this ability.

Limitations

The current study provided initial data regarding the association between SOFB and EF. We focused on a broad mapping of the EF skills most likely to be associated with SOFB, and included tasks representative of working memory, cognitive flexibility, and inhibitory control. As a result, we now have a general idea of which EFs may be most related to SOFB (i.e., working memory and cognitive flexibility). A first limitation of the current study, however, is that the specificity of these associations remains to be further narrowed. Both working memory and cognitive flexibility are multifaceted constructs (Anderson et al., 2002; Dick, 2014; Oberauer et al., 2008; Unsworth et al., 2014), and it is possible that different patterns of correlations with SOFB would be observed in a more precise assessment of these EFs.

A second limitation is the relatively small number of false belief/true belief trials included in this study. Specifically, each participant undertook four false belief trials, but only two true belief trials. As we have seen, our true belief data revealed somewhat unexpected patterns of correlation with cognitive flexibility. Further, although the observed correlations were generally stronger for the EF-SOFB relation than the EF-FOFB, these did not significantly differ from each other (all *p*s > .143). Increasing the number of trials might provide a more accurate estimate of the role of EF in belief reasoning (Bates et al., 1992). In Study 3, we doubled the number of both true and false belief trials, as a way to enhance statistical power and to improve the reliability of our findings (Baker et al., 2019).

Finally, the variance predicted by our verbal ability assessment was non-significant. While it may be that general intellectual abilities such as verbal ability no longer play a significant role in the maintenance of ToM in adulthood, a plausible alternative is that, given the minimally verbal nature of our ToM tasks, non-verbal measures of general intelligence may be more strongly implicated. We address all these limitations in Study 3.

CHAPTER 6

Study 3

Individual Differences in Second-Order False-Belief Reasoning and Executive Processes in Adults: A Further Investigation into the Specificity of Associations

In Study 2, we found evidence of a role for EF in the adult ability to infer SOFBs. In contrast with other experimental approaches that have failed to find significant correlations or have found very low correlations, we attributed our findings to experimental stimuli that were sufficiently sensitive to capture inter-individual variability. Providing age-appropriate testing conditions, as well as minimizing the risk of uniformity in performance appeared to have been fruitful in uncovering that EF relates to SOFB understanding beyond childhood. More precisely, we found that both working memory and cognitive flexibility were significant predictors of SOFB performance, independently of other controls such as FOFB performance and verbal ability. These findings provided working hypotheses of the EFs most central to SOFB performance, as well as those most involved in FOFB, and true-belief reasoning. In Study 3 (preregistered at osf.io/x5ubp), we aimed to replicate and extend our earlier findings by conducting a more in-depth assessment of the specific contribution of both working memory and cognitive flexibility to SOFB reasoning.

As mentioned earlier, working memory and cognitive flexibility are complex constructs that can each be subdivided into finer, more specific sub-processes (e.g., Mayr & Kliegl, 2000; Unsworth & Engle, 2007). Recall that in Study 2 we found a significant association between working memory and SOFB. However, the set of tasks selected to index participants' working memory skills - *N*-Back and Backward Digit Span - only allowed us to evaluate variability in manipulation/processing capacity across individuals. We suggested that one aspect of the SOFB task which could have been particularly demanding of information manipulation skills is the merging of Agent 1's beliefs with Agent 2's beliefs, at the point in the sequence where the right belief must be attributed to the right agent. Such operation arguably requires not only the tracking of a considerable mental state load, but also the coordination and manipulation of the mental state information that is being temporarily held in mind.

On the other hand, an arguably more parsimonious explanation is that the *storage* component of working memory underpins the EF-SOFB relation and that the manipulation component is less central (Bayliss et al., 2003; Engle et al., 1992; Friedman & Miyake, 2004; Unsworth et al., 2009; Süb et al., 2002). Holding in mind two mental states may be driving the association with working memory, without the need for information manipulation as such. To examine whether these sub-processes may account differently for variability in SOFB performance, participants in Study 3 undertook two working memory capacity tasks (i.e., Forward Digit Span and Corsi Blocks), in addition to the same set of working memory manipulation tasks included in Study 2.

Cognitive flexibility is also believed to consist of at least two complementary dimensions – the ability to switch from one task to another (task-switching) and the ability to maintain and retrieve various task sets (set-shifting) (Dajani & Uddin, 2015; Mayr & Kliegl, 2000). In Study 2, participants were presented with measures of task-switching only (i.e., Trail-Making and Plus/Minus). Because these tasks correlated relatively uniformly across ToM measures, whether the beliefs to be inferred were of first/second order, or true/false content, we suggested that perhaps task-switching skills were involved whenever a task made requirements in perspective-switching, regardless of the difficulty or veracity of the belief inference.

Yet, it is possible that other aspects of cognitive flexibility, such as set-shifting, may be more selectively involved in SOFB reasoning. The rationale for suggesting that set-shifting may be particularly relevant for SOFB comes in part from developmental studies, showing a systematic bias in younger children's (wrong) answers to the SOFB test question. Children who fail the second-order task do not answer at random, or according to reality, but instead continue to give first-order responses to SOFB questions. To illustrate, Arslan, Taatgen and Verbrugge (2017) presented 5-6-year-old children with SOFB scenarios of the Perner and Wimmer (1985) type, but modified in such a way that allowed for discerning three possible levels of reasoning among children's responses (i.e., zero, first, or second-order). Applied to the Ice Cream Truck story, this was achieved through the addition of a third location, the playground, where the ice cream man would drive to, unbeknownst to either main character. In this case, a zero-order answer to "Where does John think that Mary thinks that the ice cream man is?" would correspond to "the playground", while "the church" would be a first-order answer, and "the park" would be the accurate, SOFB answer. Arslan et al. found that over 65% of children's incorrect answers fell into the first-order category, while only 29% were based on zero-order reasoning. One plausible explanation is that after they have inferred John's FOFB, children fail to answer according to a new set of "second-order" rules. Within the same task, children show perseveration errors in their responses, and continue to focus on the agent's FOFB, instead of shifting to their SOFB. We tested the influence of this potentially relevant aspect of cognitive flexibility, by selecting two tasks representative of set-shifting (i.e., numerical decision and category switch), which were administered in addition to the task-switching measures presented in Study 2.

Additionally, Study 3 addresses some limitations identified in Study 2. Notably, we took two steps toward enhancing statistical power: We increased our sample size and improved measurement precision by doubling the number of true belief and false belief trials presented to participants (Baker et al., 2019; Maxwell et al., 2008). Further, our earlier research found very little evidence of shared variance between participants' verbal ability and SOFB. Here, as an alternative, we administered two measures of fluid reasoning to assess the extent to which this non-verbal aspect of intelligence significantly influences the relationship between SOFB and EF. This decision was motivated by previous reports of some linkage between fluid reasoning and both first-order ToM (Ibanez et al., 2013) and EF (e.g., Friedman et al., 2006; Fukuda et al., 2010).

Study Aims and Predictions

Broadly, the aims of Study 3 were twofold: replication and extension of our previous findings. As in Study 2, we examined how individual differences in SOFB relate to individual differences in EF. To increase the specificity of our earlier findings, we narrowed our focus to the two particular EFs that appeared to be best predictive of SOFB performance in the earlier study: working memory and cognitive flexibility. Following the findings of Study 2, we explored the following issues in Study 3:

(1) If working memory and cognitive flexibility truly play a role in the second-order ToM skills of adults, then the findings of Study 2 should replicate, with significant associations between these components of EF and performance on the SOFB task.

(2) Our prediction in the earlier research as to whether specific EFs would make unique contributions to SOFB reasoning was largely exploratory given the relative scarcity of prior

work. However, in Study 2 we found working memory to be a somewhat stronger predictor of SOFB than cognitive flexibility. If working memory is truly the dominant predictor, then it should emerge here again as the best predictor of individual SOFB performance, above and beyond cognitive flexibility and other control variables (i.e., FOFB and non-verbal ability).

(3) We tested whether working memory was related to SOFB because of the multiple demands in information transformation (manipulation), or more simply because of the arguably high mental state load to be held in mind (storage).

(4) Similarly, we tested whether the task-switching component of cognitive flexibility is more central to SOFB than the set-shifting component.

(5) As in Study 2, our findings had the potential to inform the extent to which there is continuity or discontinuity in the EFs supporting FOFB reasoning versus the EFs supporting SOFB reasoning. If similar relations with EF are found across both FOFB and SOFB, then this may imply some degree of continuity in the processes in play across these two forms of reasoning (Leslie et al., 2005; Perner et al., 2002). Alternatively, some EFs may be more specific to either FOFB or SOFB, indicative of an executive distinction between FOFB and SOFB (Apperly et al., 2009; Miller et al., 2018).

Method

Participants

Participants were 120 undergraduate students (86 females) recruited from the Victoria University of Wellington (New Zealand) research participant pool, who participated in exchange of partial course credits. Participants ranged in age from 17 to 32 years (M = 19.7 years). All participants were fluent in English and had normal or corrected-to-normal vision.

Measures

Our task-selection process was directly motivated by the goals of the current research. Building on our test battery from Study 2, we ensured comparability with our previous methodology, but also included new measures representative of two additional EF sub-components (i.e., storage and set-shifting). As in Study 2, task administration was computerized and implemented by the open-source Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014). All participants completed the task battery in the following fixed order: Mental Rotation, Pattern Comparison, FOFB, SOFB, Trail-Making, Plus/Minus, Category Switch, Numerical Magnitude Decision, Corsi Blocks, Digit Span Forward, Digit Span Backward, and *N*-Back.

Non-Verbal Ability Measures

As measures of fluid intelligence, we included two indices of participants' non-verbal reasoning: a mental rotation task (Cooper & Shepard, 1973) and a pattern comparison task (Perez et al., 1987). Mental rotation is a well-known task which consists in the simultaneous presentation of two 2D stimuli (e.g., "L" shape) that differ in axial orientation. Participants were asked to mentally compare and judge whether each pair of stimuli was identical or non-identical. Scores were calculated on the basis of the reaction time for making a judgment (64 trials). In the pattern comparison task, participants were asked to discern whether pairs of 4x4 matrices display identical dot patterns or not. This task's demands were relatively simple, and participants' reaction times served as an index of perceptual speed (60 trials).

False-Belief Measures

First-order false-belief and Second-order false belief: The procedure for these tasks was the same as in Study 2, with the exception that four true-belief trials (instead of two in Study 2) and eight false-belief trials (instead of four) were included.

Executive Function Measures

Cognitive flexibility: task-switching component

Trail Making Test: The procedure for this task was the same as in Study 2.

Plus/Minus task: The procedure for this task was the same as in Study 2.

Cognitive flexibility: set- shifting component

Category-switch: This task was a variation of Mayr and Kliegl's (2000) original category-switch task, designed to assess set-shifting abilities. The set of stimuli comprised simple shapes (i.e., circle, star, cross, square) of varying colour (i.e., green, blue, orange, yellow), and with one of six letters on them (i.e., A, Q, K, L, O, X). The procedure required identifying the shape that matched a highlighted target shape, on the basis of a particular feature indicated at the top of the screen (e.g., shape, colour, letter). Rule-shifting occurred in randomized fashion across trials. There was a total of 78 trials, and the average completion time per trial was used as an index of set -shifting ability (lower scores = better performance).

Numerical Magnitude Decision: In this task, sets included the appropriate stimulus-response mappings for a given trial, such that the task goal remained constant (e.g., press button x for stimulus y) but stimulus-response mapping reversal occurred every X number of trials (e.g., press button y for stimulus x) (Dajani & Uddin, 2015). Participants were presented with a series of target digits (between 0 and 9), which were to be classified as smaller or larger than 5. The initial set of stimulus-response mappings had the participant press the right shift key for "smaller than 5" and the left shift key for "larger than 5" (hereafter referred to as block A). Then, in subsequent blocks of trials, this response rule was reversed. Participants pressed the left shift key for "larger than 5", and the right shift key for "smaller than 5" (block B). Two "A" blocks (20 trials each), and two "B" blocks (20 trials each) were completed in the following order: A-B-B-A. Each block consisted of 32 trials, and a shifting cost was obtained by subtracting the RT to blocks between which set-shifting was required (A-B and B-A), from RT to blocks between which no set-shifting was required (B-B). Low shifting costs therefore reflected better set-shifting efficiency.

Working memory: storage component

Corsi blocks: This task presented a display consisting of a set of nine square targets arranged irregularly on the screen. Starting with sequences of two squares, the squares flashed one at a time in random sequences of increasing length. Participants were asked to recall and reproduce the observed sequences by using a mouse click on the squares in the correct order. The number of steps in a sequence increased by one as participants succeeded at two consecutive trials, for a maximum of nine steps (Corsi, 1972). Following Miyake et al. (2000b), this task provided an index of working memory storage capacity based on the longest sequence correctly reproduced.

Digit Span (Forward): In this task, participants were presented with sets of increasingly longer series of digits (e.g., "3-2-1-4") and asked to recall the series in the same order as presented (e.g., "3-2-1-4"). Presentation begins with three digits in a series, and the number of digits increased by one as participants succeeded at two consecutive trials for a maximum of ten digits per series. The longest series of digits correctly recalled indicated one's working memory storage capacity (Hebb, 1961).

Working memory: manipulation component

Digit Span (Backward): The procedure for this task was the same as in Study 2.

N-Back: The procedure for this task was the same as in Study 2.

Results

The following analyses conformed to our preregistered plan, and were all performed in *R* (R core team, 2019). In keeping with our previous analytic procedure, all task distributions were examined for departure from normality. Skewness and kurtosis of four reaction time-based measures (Pattern Comparison, SOFB, FOFB, TMT) were improved with Yeo-Johnson transformations prior to estimating the correlations (Yeo & Johnson, 2000). To minimize the impact of outlying responses, any score beyond 1.5 times the maximum and/or minimum values of the interquartile range ($Q3$–$Q1$) were winsorized to the nearest non-outlying value (Tabachnick & Fidell, 1996). This procedure was applied to each individual task distribution, and resulted in the replacement of 2.05 % of the overall data. Finally, where higher values of the dependent variables indicated better performance (*N*-Back, Digit Span, and Corsi Blocks), scores were reversed to ensure consistency in scoring interpretation across all analyses. As in study 2, we analyzed reaction time data for correct responses only (across all false-belief and EF tasks).

Descriptive statistics for all measures are summarized in Table 13. As can be seen, all task distributions exhibited adequate variability in performance and met the requirements for approximate normality (i.e., skewness < 2 and kurtosis < 4).

Table 13

Descriptive Statistics for all Measures

	Measure	DV	Mean	*SD*	Min	Max	Skew	Kurt
ToM	FOFB	RT	740.14	570.17	181.29	3424.67	.01	-.10
	FOTB	RT	736.07	537.66	184.38	2356.00	.39	-.23
	SOFB	RT	956.67	858.12	250.88	6394.50	.06	-.63
	SOTB	RT	853.51	599.20	217.25	3362.00	.05	-.85
Non- Verbal Ability	Pattern Comparison	RT	1481.56	267.65	828.90	2111.59	-.10	-.72
	Mental Rotation	RT	1093.16	206.34	734.02	1548.00	.50	-.42
Task-Switching	Plus/Minus	Shifting effect	493.92	554.70	-993.89	1742.00	-.06	-.40
	TMT	Shifting effect	305.12	514.91	-2324.23	1055.11	.63	-.34
Set-Shifting	Num Magnitude	Shifting effect	111.50	106.69	-151.15	506.65	.20	-.34
	Category Switch	Shifting effect	2364.05	459.62	1458.89	3432.74	.68	-.23
Storage	Corsi Blocks	Span length	6.60	1.42	2	9	-.95	2.04
	DSpan Forward	Span length	5.88	1.30	2	9	-.29	.33
Manipulation	*N*-Back	Number of CR	56.99	4.65	46	64	-.62	-.26
	DSpan Backward	Span length	4.92	1.59	2	10	.17	-.30

Note. ToM = Theory of Mind; FOFB = First-Order False-belief; FOTB = First-Order True-belief; SOFB = Second-Order False-Belief, SOTB = Second-Order True-Belief; CR = Correct Responses; RT= reaction time in milliseconds; TMT = Trail Making Test; Num Magnitude = Numerical Magnitude Decision; DSpan = Digit Span; Skew = Skewness; Kurt = Kurtosis; *N* = 120 except for FOFB and SOFB: *N* = 119 as one participant failed all trials on these tasks.

Correlations within and across EF and Non-Verbal Ability Assessments

In line with our primary goal for this research, we began by decomposing our test battery into four different sub-components, represented by specific EF abilities: working memory storage, working memory manipulation, set-shifting, and task-switching. As pairwise analyses indicated that performance tended to covary within each sub-component (Table 14), four composite scores were devised for each participant: scores on the Forward Digit Span/Corsi Blocks, and Backward Digit Span/*N*-Back were respectively averaged into working memory storage and working memory manipulation composite, while scores on the Numerical Magnitude Decision/Category Switch and Plus-Minus/TMT were respectively averaged into a set-shifting and a task-switching composite. Since construct intercorrelations were also observed, and in keeping with our Study 2 replication goal, three additional composite estimates were derived from the averaging of storage/manipulation scores (working memory), set-shifting/task-switching scores (cognitive flexibility), and all EF measures (EF). Finally, given the significant correlation between Pattern Comparison and Mental Rotation, data were collapsed into a non-verbal ability composite score. Zero-order correlations were calculated among all measures and are displayed in Table 14.

Table 14

Correlation Matrix for all Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. FOFB	-												
2. FOTB	.55***	-											
3. SOFB	.37***	.41***	-										
4. SOTB	.52***	.52***	.36***	-									
5. Patt Comparison	.11	.21*	.14	.28**	-								
6. Mental Rotation	.24*	.20*	.28**	.07	**.32*****	-							
7. Plus/Minus effect	.19*	.12	.21*	.10	.04	.19*	-						
8. TMT effect	.02	.21*	.15	.09	.07	.03	**.21***	-					
9. Num Magnitude	.27**	.19*	.17	.14	.12	.24**	.32***	.04	-				
10. Category Switch	.09	.15	.15	.09	.02	.21*	.42***	.10	**.18***	-			
11. Corsi Blocks	.13	.18	.10	.08	.08	.16	.09	.21*	.12	.18	-		
12. DSpan Forward	.16	.03	.06	.04	.05	.19	.01	.09	.02	.33***	**.24****	-	
13. *N*-Back	.15	.10	.22*	.08	.13	.20*	.26**	.07	.18	.37***	.27**	.24**	-
14. DSpan Backward	.16	.13	.26**	.03	.09	.27**	.37***	.01	.27**	.39***	.39***	.41***	**.38*****

Note. FOFB = First-Order False-Belief; FOTB = First-Order True-Belief; SOFB = Second-Order False-Belief; SOTB = Second-Order True-Belief; Patt Comparison = Pattern Comparison; TMT = Trail-Making Test; Num Magnitude = Numerical Magnitude Decision; DSpan = Digit Span. Cells denoting pairwise correlations within a EF sub-domain are bolded. * $p < .05$; ** $p < .01$; *** $p < .001$.

Assessed next was whether there were intercorrelations among these EF composite scores, and whether EF was associated with non-verbal ability. As Table 15 shows, scores were significantly intercorrelated within (storage and manipulation, $p < .001$, set-shifting and task-switching, $p < .001$) and across all EF indices (working memory and cognitive flexibility, $p < .001$). Non-verbal ability, however, did not significantly correlate with any of the EF scores (all *p*s > .092).

Table 15

Zero-Order Correlations among Composite Measures of EF and Non-Verbal Ability

Measure	1	2	3	4	5	6	7
1. Non-Verbal Ability	-						
2. Working memory	.08	-					
3. Storage	.16	$.67^{***}$	-				
4. Manipulation	.04	$.95^{***}$	$.43^{***}$	-			
5. Cognitive flexibility	.16	$.47^{***}$	$.31^{**}$	$.45^{***}$	-		
6. Set-Shifting	.11	$.47^{***}$	$.32^{**}$	$.44^{***}$	$.83^{***}$	-	
7. Task-Switching	.15	$.34^{***}$	.21	$.33^{***}$	$.87^{***}$	$.45^{***}$	-

* $p < .05$; ** $p < .01$; *** $p < .001$.

False-Belief Assessment

We next addressed the main question of interest: What is the relative contribution of each EF construct to SOFB reasoning? We examined this question with correlational analyses conducted for each construct (see Table 16). Significant associations were obtained between SOFB and composite EF ($p < .018$), driven more or less equally by both composite working memory ($p < .006$) and composite cognitive flexibility ($p < .018$). Finer-grained analysis revealed that only the manipulation component of working memory, and not storage, contributed significantly to the working memory-SOFB association ($p < .004$). Similarly, task-switching was the sole cognitive flexibility contributor to significant variance in SOFB ($p < .020$). Upon controlling for non-verbal ability, however, only working memory manipulation held up as a sub-component significantly predictive of variance in SOFB ($p < .004$). As shown in Table 16, SOFB was also significantly related to FOFB ($p < .001$), even with non-verbal ability held constant ($p < .001$). The previous set of analyses were thus reconducted, this time accounting for variance in FOFB. Comparable outcomes were obtained from this analysis; working memory manipulation continued to share significant variance with SOFB ($p < .013$).

Table 16

Zero-Order and Partial Correlations between EF, Non-Verbal Ability, SOFB and FOFB Reasoning

	SOFB	FOFB
EF composite	$.22^{*}$/.18/.16	$.19^{*}$/.17
Working memory	$.25^{**}$/$.24^{**}$/$.20^{*}$	$.20^{*}$/$.20^{*}$
Storage	.10/.06/.04	$.18^{*}$/.17
Manipulation	$.27^{**}$/$.26^{**}$/$.23^{*}$	.17/.17
Cognitive flexibility	$.22^{*}$/.18/.16	$.19^{*}$/.17
Set-Shifting	.15/.13/.11	.14/.13
Task-Switching	$.21^{*}$/.17/.15	$.19^{*}$/.16
Non-Verbal Ability	$.24^{**}$	$.19^{*}$
FOFB	$.37^{***}$/$.36^{***}$	-

Note. EF = Executive Functioning; FOFB = First-Order False-Belief. This table shows raw correlations/partial correlations controlling for non-verbal ability/partial correlations controlling for non-verbal ability and FOFB performance.

* $p < .05$; ** $p < .01$; *** $p < .001$.

We then ran similar analyses on FOFB data to establish the similarities and differences in the false-belief-EF associations across levels of false-belief reasoning complexity (Table 16). Overall, this analysis showed that FOFB was significantly associated with composite EF ($p <$

.037), as well as composite working memory ($p < .026$) and composite cognitive flexibility ($p < .036$). Of the two cognitive flexibility components assessed, task-switching emerged here again as selectively predictive of FOFB ($p < .044$), but unlike SOFB, storage ($p < .046$) was the working memory component driving the association between FOFB and composite working memory. Controlling for non-verbal ability, however, revealed that only the working memory composite remained associated with FOFB ($p < .029$).

Hierarchical Regression Analysis

To assess whether EF exerts independent predictive value, two hierarchical multiple regression models were compared with different sets of predictors and SOFB as the outcome variable. These models are summarized in Table 17. Model 1 assessed how control variables alone (Non-Verbal Ability and FOFB) predicted SOFB. In this model, both FOFB and Non-Verbal Ability accounted for significant variance in SOFB performance (16.43%). The addition of the EF variable that had significantly predicted SOFB in univariate analyses (i.e., working memory manipulation; Model 2) increased the proportion of shared variance to 20.01%, and, importantly, all three variables contributed unique predictive value to SOFB.

Table 17

Hierarchical Regression Models Including all Variables Significantly Predicting SOFB

	Regression Coefficients				
Predictors	*B*	*SE*	*t*	ΔR^2	*F*
Model 1				.164	12.60^{***}
Non-verbal ability	< .001	< .001	2.345^{*}		
FOFB	.348	.085	4.099^{***}		
Model 2				.201	10.88^{***}
Non-verbal ability	< .001	< .001	2.338^{*}		
FOFB	.317	.084	3.774^{***}		
Working memory manipulation	.076	.030	2.510^{*}		

Note. FOFB = First-Order False-Belief.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Exploratory Analysis of the Relations between True-Belief and EF

As in Study 2, we also examined the contribution of EF to true-belief reasoning. As Table 18 shows, SOTB was only correlated with FOTB ($p < .001$), and not with EF (all *p*s > .058). With the exception of non-verbal ability ($p < .013$), no further significant correlations emerged from similar analyses focusing on FOTB data (all *p*s > .065). Together, these results indicate that true-belief reasoning may only recruit EF to a very limited extent (Table 18).

Table 18

Exploratory Correlation Analysis between EF, SOTB and FOTB Reasoning

	SOTB	FOTB
EF composite	.12/.09/.04	.16/.13
Working memory	.08/.07/.01	.15/.13
Storage	.08/.06/.01	.14/.10
Manipulation	.07/.07/.01	.13/.12
Cognitive flexibility	.12/.09/.04	.16/.14
Set-shifting	.09/.08/.01	.17/.15
Task-switching	.10/.08/.05	.12/.08
Non-verbal ability	.15	$.23^{*}$
FOTB	$.52^{***}/.51^{***}$	-

Note. EF = Executive Functioning; FOTB = First-Order True-Belief. This table shows raw correlations/partial correlations controlling for non-verbal ability/ partial correlations controlling for non-verbal ability and FOTB performance.

* *p* < .05; ** *p* < .01; *** *p* < .001.

Paralleling the findings of Study 2, SOTB appears to be almost uniquely predicted by variance in FOTB understanding. This reinforces our earlier suggestion that EF may be restrictively recruited by belief reasoning operations that involve thinking about an agent's false-belief. We nonetheless conducted, as a final analysis, a hierarchical regression predicting SOFB from working memory manipulation, non-verbal ability, FOFB, and FOTB/SOTB (Table 19).

Table 19

Hierarchical Regression Model Including all Variables Significantly Predicting SOFB

	Regression Coefficients				
Predictors	*B*	*SE*	*t*	ΔR^2	*F*
				.265	9.513^{***}
Non-verbal ability	< .001	< .001	1.75		
FOFB	.184	.092	1.99^{*}		
FOTB	.182	.113	1.61		
SOTB	.202	.090	2.23^{*}		
Working memory manipulation	.066	.029	2.25^{*}		

Note. FOFB = First-Order False-Belief; FOTB = First-Order True-Belief; SOTB = Second-Order True-Belief

* $p < .05$; ** $p < .01$; *** $p < .001$.

All indices but non-verbal ability and FOTB accounted for significant variance in SOFB performance. Importantly, even in this conservative analysis, working memory manipulation contributed to SOFB above and beyond the variance accounted for by true-belief reasoning.

Discussion

In this study we examined the specificity of the relations between the adult capacity for SOFB and EF. Our previous findings had indicated that some common skills may underlie these two constructs, especially within the domains of working memory and cognitive flexibility. A focus on more precise abilities was adopted in the current study. Here, as in Study 2, participants saw SOFB-inducing stimuli with particular aspects of the design manipulated to increase the sensitivity, but decrease the peripheral demands of traditional tasks. We assessed participants' responses in this task in relation to two domains of executive performance examined previously, subdivided into finer areas.

Summary of Experimental Findings

First, the construct of working memory was differentiated into storage (holding information in mind) and manipulation (mental processing/transformation of this information). In Study 2 we had found a low-to-moderate association between variability in participants' manipulation capacity and SOFB. We sought to reproduce this in Study 3, in addition to gauging the effect of the more parsimonious ability for information storage. We found that, in line with our previous observations, participants' skills for information manipulation significantly correlated with their second-order, but not first-order FB performance. Further, for SOFB, this correlation held up above and beyond the influence of all other potential correlates (i.e., non-verbal ability, FOFB, FOTB/SOTB). In contrast, the storage component of working memory contributed very little to the explanation of SOFB performance, and only correlated with FOFB before controlling for non-verbal ability.

Second, the specific patterns of correlations between cognitive flexibility and SOFB observed here showed only partial replication of our prior findings. In addition to the ability to perform different tasks on a single set of stimuli (task-switching) assessed in Study 2, participants' cognitive flexibility was also indexed by their ability to shift from one mental set to another (set-shifting). Though initially replicated in raw correlations, the association between SOFB and task-switching did not appear to be especially robust, and did not hold up over non-verbal ability. A similar pattern was observed for FOFB, which correlated with task-switching, but not after controlling for non-verbal ability. Contrary to our suggestion that set-shifting may be a construct related to SOFB, the addition of set-shifting measures did not reveal any new patterns of associations with SOFB or FOFB. Further, Study 2 had revealed somewhat unexpected correlations between cognitive flexibility and true-belief performance. These were not replicated in Study 3. The relation between these constructs found in Study 2 may thus have been spurious or else the relation is a very weak one that is not reliably detected.

Limitation

In Study 2 we noted that a considerable caveat was the fact that the correlations between SOFB and EF were not significantly different from the correlations between FOFB and EF. This caveat is important to reiterate here, as we observed no significant differences between any of the EF-SOFB and EF-FOFB pairwise correlations (all *p*s > .212). Additionally, even though we increased the specificity of our EF measures, the number of tasks representative of each key EF remained relatively low. Increasing the size of the test battery may be a further way to increase statistical power and improve the reliability of our findings (Miyake et al., 2000a).

CHAPTER 7

General Discussion

The research undertaken in this thesis was designed to provide novel insights into the linkage between ToM and EF. We were particularly interested in exploring the possibility of links between EF and SOFB reasoning, an aspect of ToM whose cognitive underpinnings are largely understudied. Within that interest our central focus was on whether these links might be found for adults and how specific those links might be. Along the way we also drew on, and empirically explored relations between EF and ToM in children, as a point of comparison to adults. Our approach was systematic: We began by conducting a meta-analytic review of the extant literature on the topic in both adults and children (Study 1); we then examined some of the many unanswered questions about EF and ToM through a broad assessment of the links between EF and SOFB in adults (Study 2); and we followed up this general assessment with an analysis focused on finer-grained components of EF (Study 3). This final chapter summarizes the main findings from each study, and considers the implications of the findings for our broader understanding of the relations between EF and ToM. Finally, we present some suggestions for potential ways forward to advance our understanding of the extent to which these two fundamental aspects of human cognition are intertwined.

Relations between SOFB and EF in Children

In Study 1 we used meta-analytic methods to systematize the prior literature on the association between EF and SOFB reasoning in both adults and children. We incorporated comparative evidence from school-age children to gauge the relations between SOFB and EF in an age group in which these constructs have been heavily studied.

One of the main conclusions from this meta-analysis was that, for school-age children, EF is significantly linked to individual differences in SOFB understanding. The pooled results for this age group showed a moderate positive association between the two constructs (r = .25). The strength of this association resisted statistical adjustment for age, and did not differ on the basis of children's cultural background (i.e., Eastern or Western). Further, the available data allowed for the comparison of four EFs (attention, working memory, inhibitory control, and cognitive flexibility), which all contributed significantly and relatively equally to SOFB performance.

Earlier, we discussed two kinds of explanations - expression and emergence - that represent the dominant ways of theorizing about the EF-ToM relation in children. The expression account appeals to executive limitations that preclude the accurate expression of already present false-belief understanding. The emergence account posits functional links between EF and ToM, such that progress in specific EFs provides some of the foundational skills for newly emerging false-belief understanding. A certain level of continuity across tasks of varying complexity is predicted by proponents of an emerging role for EF in preschool ToM. More precisely, the prediction is that if some EF skills are prerequisite to the very emergence of some ToM concepts, then EF should correlate with tasks assessing all of these ToM concepts, regardless of their relative executive demands (Carlson et al., 2015). Expression accounts, on the other hand, predict some degree of dissociation between ToM measures imposing lower versus higher executive demands. An effect of executive load should be observed in the relation between EF and ToM, such that performance on ToM tasks imposing higher executive demands should correlate more strongly with EF (Perner et al., 2002).

Our meta-analytic findings evidenced some cross-cultural stability in the relations between children's SOFB and EF, which may speak to these developmental accounts. Cross-cultural differences have never been formally assessed in the second-order ToM domain, but our meta-analytic findings are consistent with cross-cultural research on children's first-order ToM and EF. Children of some Eastern cultures in which self-regulation is emphasized much earlier in development (e.g., China, Korea) often outperform Western children on measures of EF, but tend to display poorer or at least no better ToM skills. Yet, these cultural differences have very little influence on the relation of FOFB to EF, which is consistently found across Eastern and Western societies alike (Duh et al., 2016; Oh & Lewis, 2008; Sabbagh et al., 2006; Wang et al., 2016). These findings, along with our meta-analytic findings, further suggest that the expression of SOFB is not affected by culture-related advantages in EF. Instead, the executive correlates of SOFB appear to be consistent across Western and Eastern societies, and explain individual differences to a similar extent, relatively independently of children's more or less advanced EF skills. Together, these cross-cultural findings align with an emergence hypothesis suggesting that EF skills may be necessary, but not sufficient for children to develop a deeper understanding of mental states (Moses, 2005).

Lastly, recall that an earlier meta-analysis (Devine & Hughes, 2014) found uniformity in the association between EF and both FOFB and SOFB across studies: Even though presumably more demanding, SOFB did not correlate more strongly with EF. As suggested by the authors, this finding could challenge the idea that when task demands are high, the expression of false-belief understanding should be significantly more dependent on EF (Perner et al., 2002), and instead fit within an emergence account predicting similar strengths of association across ToM tasks of varying executive demands (Carlson et al., 2015). A caveat, however, was that some of the effect sizes reported in Devine and Hughes were calculated on the basis of composite scores, also including children's FOFB performance. There was thus some uncertainty around the specific role of SOFB in driving these associations, which makes it harder to firmly conclude whether EF relates to SOFB more or less strongly than it relates to FOFB. Similarly, although our meta-analytic findings revealed an overall effect size of smaller magnitude (r = .25) than that typically reported in the first-order ToM literature (r = .38; Devine & Hughes, 2014), the children sampled here were generally older, and presumably had greater EF skills (possibly even beyond the skills needed for both emergence and expression). The evidence presented here may thus not be able to speak strongly for or against either account.

Relations between SOFB and EF in Adults

Although the meta-analysis of the studies conducted in school-age children confirmed significant links between SOFB and EF, our analyses of adult research evidenced several factors limiting the interpretability of the pooled results for this age group. We found that EF was related to SOFB only to a small and non-significant extent (r = .09). However, the presence of factors such as narrow test batteries, non-random sampling, and non-age-adapted measures posed several constraints on the interpretation of the data. Adult studies not only accounted for a very small percentage of the meta-analyzed literature, but the majority of these studies only assessed the relation between EF and SOFB to a limited extent. For instance, most studies administered a single EF or SOFB assessment, and presented adults with a SOFB task originally designed for 5-6 year-old children. Additionally, effect sizes from all but one study consisted of control group data included in the context of clinical research with the primary goal of assessing, not the relations between SOFB and EF, but the more general cognitive functioning of a given patient group. Together, these limitations may well have masked important variability in performance, and constrained the conclusions that could be drawn. Moreover, the positive results obtained from developmental research were a preliminary indication that beyond the emergence of FOFB, EF plays a lasting role in the new skills that individuals add to their ToM repertoire. Combined with evidence of heterogeneity across individuals' motivations and skills for second-order ToM (de Weerd et al., 2017; Kinderman et al., 1998), the child meta-analysis provided a starting point to hypothesize that, if tested with methods of increased sensitivity, the influence of EF on SOFB might be found to persist through adulthood.

To test this hypothesis, two follow-up studies were carried out. Our approach in Study 2 consisted of a broad assessment of adults' SOFB skills in relation to performance on three dimensions of EF adopted by most dominant models in the adult literature: inhibitory control, working memory, and cognitive flexibility (Friedman et al., 2006; Miyake et al., 2000b). Participants undertook a novel SOFB task, devised for the threefold purpose of providing a new, age-appropriate methodology, while reducing the demands of the classic task and creating a structure more readily comparable to a FOFB task. This new task appeared to be sensitive enough to capture meaningful variability across participants, as the outcome of our analysis revealed positive associations between SOFB and global EF ($r = .39$). In addition, and more specifically, we found that both working memory and cognitive flexibility correlated with SOFB performance ($r = .31$ and $r = .40$, respectively), independently of variance accounted for by other controls such as FOFB performance and verbal ability. Inhibitory control, on the other hand, despite being a well-established correlate of children's first-order ToM, shared very little variance with adults' SOFB performance ($r = .11$).

Study 3 tested the robustness of these findings, via a design in which participants' EF skills were assessed more finely. We focused on the specific EFs that had been shown to be most predictive of SOFB performance in Study 2 (working memory and cognitive flexibility) and explored which of their respective sub-components more precisely contributed to the associations observed previously. Accordingly, the construct of working memory was broken into two components: In addition to the manipulation component assessed in Study 2 (mental processing/transformation of information), we also assessed participants' storage capacity (holding information in mind). Likewise, cognitive flexibility was subdivided into two components: Task-switching (performing different tasks on a single set of stimuli), as in Study 2, and set-shifting (shifting from one mental set to another). In line with our previous observations, our analysis revealed a positive association between SOFB and global EF (r = .22). Further, participants' skills for information manipulation in working memory, but not storage, significantly correlated with their SOFB performance (r = .27), as did their task-switching abilities (r = .21), but not set-shifting. Upon controlling for participants' non-verbal ability and FOFB performance, however, only the information manipulation component of working memory contributed to SOFB beyond the variance accounted for by these former factors (r = .23).

The persistent correlation with working memory across two studies raises interesting questions regarding the nature of SOFB reasoning. We suggested earlier that at least two main features of the task could explain why working memory appears to be particularly important for adults' SOFB, but not necessarily for their FOFB performance. First, we proposed a mental state load explanation, whereby the requirement to hold in mind two, versus only one mental state, is what could be driving the correlation between SOFB and the storage component of working memory rather than the manipulation component. Given that our SOFB task was designed to match the FOFB task on all inferential aspects (i.e., belief retention period, number of characters and belief-inducing events, etc.), the number of beliefs specific to each task was one of the only remaining factors potentially enhancing the working memory demands of the SOFB task. It is possible then, that if SOFB reasoning does impose higher working memory demands than FOFB reasoning, not because of extraneous task factors but because of differences in mental state load, then perhaps the difference between the two types of false-belief inferences is best understood in terms of complexity-related factors. There may be a quantitative distinction between their respective working memory loads, such that SOFB could be considered as a more taxing, but not necessarily different form of belief reasoning than FOFB (Apperly, 2012; Baron-Cohen et al., 1997; Leslie et al., 2005; Lockl & Schneider, 2007; Sullivan et al.,1994).

Second, we also discussed the idea that one of the elements separating second-order ToM from its first-order counterpart is the recursive structure of the task (Miller, 2009, 2012). Specifically, the mental states of Agent 1, at one point in the sequence, must be integrated within the mental states of Agent 2, which allows reasoning through the interactions that link these mental states together. Such integration is presumably demanding of working memory, not only because of the amount of information to be maintained in mind, but also because this information must be kept active long enough to be mentally manipulated and transformed. The absence of such process in FOFB reasoning may be taken to suggest some type of qualitative distinction in the skills underlying SOFB understanding (Austin et al., 2014; Bock et al., 2015; Farrant et al., 2014; Lagattuta et al., 2015; Miller et al., 2018; Perner et al., 2002).

Across both Studies 2 and 3, the ability to process information in working memory, as opposed to strictly storing this information, was most related to performance on the SOFB task. That finding is more consistent with the recursive hypothesis just outlined than with a complexity hypothesis based on the need to hold two mental states in mind: What drives the association between working memory and SOFB may be the need to reason through the interaction of beliefs held by others, via a mental manipulation of the relevant information (Braüner et al., 2016).

Relations Between FOFB, SOFB, and EF

We have previously explained how the adult ability to use ToM recursively has not been the topic of many empirical investigations. Not only has SOFB been neglected in terms of individual difference and EF studies, but the task itself as well as its general components have also been largely under-examined. Prior to the studies reported here, we did not know, for instance, which components of the task make it so challenging, and whether the task departs in some meaningful ways from FOFB, or can more simply be considered as a more complex reiteration of the same thought process. To begin to answer these questions, we suggested that comparing the EFs most predictive of SOFB with those most predictive of FOFB could inform the degree to which there are executive similarities or dissociations between the two constructs.

Accordingly, in addition to assessing the executive correlates of SOFB, we examined how FOFB related to EF. Across studies, we found that there was a lack of clear associations between FOFB and any of the EF measures. The data showed only a weak to non-existent influence of the storage functions of working memory in participants' FOFB performance, and no support for an influence of working memory manipulation or any of the cognitive flexibility constructs. These findings could be argued to reflect a dissociation between FOFB and EF - which would be largely conflicting with the view that first-order ToM operates under certain executive constraints (e.g., Apperly et al., 2006; Back & Apperly, 2010), and would contradict a substantial body of evidence showing links between adults' FOFB and EF (e.g., Bradford et al., 2015; Lin et al., 2010; Wardlow, 2013). The findings would be in some ways consistent, however, with considerations of false-belief reasoning as a decomposable sequence of operations (e.g., belief encoding/maintenance/use), each drawing more or less heavily on EF (Apperly et al., 2009).

For instance, participants in the current study were instructed to simply indicate where the agent thought the object was located. For adults, this equates to a relatively effortless belief-encoding operation that may in some instances proceed independently from central or executive resources (Apperly et al., 2006; Back & Apperly, 2010; Cohen & German, 2010; Qureshi et al., 2020). As such, participants were not required to *use* this information to make action predictions (e.g., the agent will go to location X to retrieve the object), or to interpret an agent's behaviour (e.g., the agent went to location X because that is where they believed the object to be located), which are both undertaken much more deliberately, and often accompanied by a great amount of cognitive effort (Epley et al., 2004; Lin et al., 2010; Keysar et al., 2003). It is perhaps likely, then, that it is when observing these latter kinds of operations that significant FOFB-EF correlations would emerge more clearly, and that perhaps our task failed to elicit some of the operations which more reliably generate executive demands.

That said, we cannot rule out the possibility that, for adults, EF may be more selectively recruited by more complex belief inferences such as SOFB. In contrast to SOFB, FOFB may be much less demanding of executive resources, and well within the EF capacities of most adults. Hence, although FOFB requires executive skills, it remains possible that individual differences in those skills in adults are beyond the level required to succeed.

Limitations and Future Directions

Interpretation of the EF-SOFB Relation

The interpretation of the findings for adults was made difficult in part by the fact that much of the research so far has revolved around the importance of EF for the initial acquisition of FOFB understanding, and that theories of children's developing ToM do not easily extend to the more mature cognitive skills of adults. Adults have more EF resources at their disposal, and accordingly regulate their thoughts and behaviour more efficiently than children (Brocki & Bohlin, 2004; Davidson et al., 2006; Huizinga et al., 2006; Lehto et al., 2003). Likewise, adults have much more experience assessing the mental lives of others. It seems quite possible then that adults track the beliefs of others in somewhat different ways than young children, and employ executive resources differently in doing so.

Developmental theories may be suitable for certain aspects of adult cognition, but may not easily accommodate the full range of later acquired concepts. The emergence account, for instance, argues that advances in EF play an initial role in the very foundation of mental state understanding. However, these predictions are specific to the childhood period. Once children have acquired the necessary skills for belief reasoning, whether first- or second-order, it is unclear what role EF would play from an emergence perspective across wider age-ranges such as adults (Carlson et al., 2015). On the other hand, although the emergence of ToM abilities is no longer an issue in adulthood, the expression of these abilities is still at least partly reliant on EF. Adults are undoubtedly able to infer others' mental states, but do not always do so automatically, and need executive resources to deliberately step into someone else's shoes (Bernstein et al., 2011; Birch & Bloom, 2004, 2007; Epley et al. 2004; Keysar et al., 2003; Lagattuta et al., 2014; Sommerville et al., 2013).

With respect to the current studies, our findings in some ways also align with the expression account. We found that EF generally correlated more reliably and more strongly with SOFB performance, relative to FOFB performance, suggesting that the more complex a ToM task is, the more likely adults will need to employ EF resources to express their ToM understanding. It is important to reiterate, however, that the correlations between SOFB and EF (global) were not significantly greater than the correlations between FOFB and EF (p = .217 for Study 2, and p = .423 for Study 3). For the reasons mentioned above, it may be that the executive demands of SOFB are indeed higher, but may not inherently exceed the executive demands of FOFB when measured in individuals with advanced EF skills. Further research will be needed to confirm this more firmly, but this observation may cast doubts on the extent to which the EF-ToM correlation truly depends on the relative difficulties of the tasks.

Irrespective of that, a major caveat is that correlation designs do not of course inform the extent to which two variables are causally linked, and the causal direction between SOFB and EF might thus run either way. In the first-order ToM literature, most of the data are consistent with the idea that EF causally affects the development of false-belief understanding. Longitudinal investigations, for instance, typically show that early EF is a robust predictor of later ToM reasoning in childhood (Carlson et al., 2004; Hughes, 1998; Hughes & Ensor, 2007; Marcovitch et al., 2015; Müller et al., 2012; Razza & Blair, 2009), whereas early ToM is a weaker predictor of later EF. The possibility remains, however, that in the case of further ToM advances, learning to think more deeply about others' mental states may promote EF. As suggested by Perner (1988), an important aspect of second-order mental states is that they enable intentional social interaction to occur. In contrast with their first-order counterpart, these recursive states allow for reciprocity in mindreading, in the sense that when making inferences about other minds, one understands that their own mind might be the object of someone else's inferences. This may lead to greater awareness of how one is perceived in the eyes of others, and generate greater control over one's thoughts and actions. An alternative interpretation of our findings, then, could be that good SOFB skills are necessary for successful EF performance. However, our findings most strongly implicated working memory, and it is intuitively difficult to see how SOFB could generate advances in this particular EF. Nonetheless, the possibility of SOFB affecting EF remains to be evaluated empirically.

Correlational data are also open to "third-variable" interpretations, such that other variables might co-occur with the EF-SOFB relations observed here without a fundamental connection existing between these constructs. By controlling for participants' verbal and non-verbal abilities, as well as designing a SOFB task closely matched to the FOFB task, we have ruled out at least some potential confounds that might drive the relations between EF and SOFB. Nonetheless, additional factors still need to be ruled out. In line with Perner's (1991) assertation, metacognition (i.e., knowledge about cognition) may interact with both EF and ToM as all three involve some type of monitoring and regulation of one's own mental activities. One study of children has indeed found a significant association between meta-memory (i.e., knowledge about memory) and both first- and second-order ToM (Lockl & Schneider, 2007), suggesting that measures of metacognitive abilities may be a relevant addition to future research looking at the SOFB-EF relation.

Further, we did not record participants' socioeconomic status, which has been found to influence adults' individual differences in both EF (e.g., Last et al., 2018) and ToM (e.g., Germine et al., 2015). In addition, we did not control for ethnic status which might conceivably be of particular importance in a New Zealand context. Hence, recruiting a sample more representative of the general population in terms of socioeconomic status and ethnicity could rule out additional extraneous factors that may have influenced the correlations between SOFB and EF.

Specificity of the EF-SOFB Relation

Although our findings do suggest a meaningful relation between SOFB and working memory, at least two outstanding issues remain to be addressed concerning the specificity of this relation. First, we have used FOFB as a comparison point, as is often the case in developmental research. However, given that FOFB is lower than SOFB in the mental state load it imposes, *and* that it does not make any requirements in mental state embedding, then either of these factors could reasonably explain the different associations with EF. Relatedly, while we have broken down EF into some of its main sub-components to better understand which aspects of those EFs contribute most to SOFB, we have not similarly broken down the SOFB task in its sub-components to examine which of them are most predictive of the relation with EF.

With this caveat in mind, a sensible next step might be to test exactly which aspects of the SOFB task are linked to EF. Related to the consideration of false-belief reasoning as a decomposable sequence of operations explained previously - i.e., belief encoding/ maintenance/ use (Apperly et al., 2009) - SOFB similarly contains a number of different processes that may need to be studied independently. For instance, researchers could examine the extent to which inferring and holding in mind two *separate* beliefs, as opposed to two *recursive* beliefs, relates to EF. Some past studies have made related attempts by comparing the ability to comprehend different levels of recursive mental states, to the ability to remember varying numbers of facts about the stories in which the said mental states were depicted (Kinderman et al., 1998; Lewis et al., 2017). Mean error rates were disproportionately higher for the mentalizing questions, indicating that there may be something especially challenging about recursive mental state reasoning, in comparison to remembering separate pieces of factual information. However, no study so far has compared the ability to infer recursive mental states to the ability to infer a varying number of mental states with no relationship of dependence.

Designing a study that allows for separate analysis of these types of inferences in relation to EF could also help inform whether working memory is necessary to handle the mental state load of the task, or the recursive requirements of the task. Perhaps one could break the SOFB task into two sub-tasks; one looking at the ability to infer two beliefs that are not yet linked, such as Agent 1's *and* Agent 2's respective beliefs about the location of the object, and a second one looking at the ability to infer two recursive beliefs, such as Agent 1's belief about agent 2's belief. Comparing performance on each of these task components to performance on EF tasks could be a way to delineate the role of recursion in driving these associations. Given that both components would involve an equal mental state load, but only one would involve an element of recursion, then perhaps this would be an informative way to begin understanding which components of the SOFB task are most demanding of EF. If the recursive hypothesis is correct, EF should correlate with the task involving dependent mental states to a larger extent than the task without this element of dependency. Alternatively, if the mental state load hypothesis is right, then there should be no significant difference in the extent to which EF correlates with the recursive and non-recursive tasks (as both involve the same number of mental states).

An additional outstanding question is the degree to which EF is specifically involved in SOFB reasoning or extends to other types of recursive mental states. In the first-order ToM literature, EF has been found to correlate with false-belief specifically, but not closely matched tasks assessing children's understanding of desires and pretense (e.g., Moses et al., 2003). With respect to the current studies, we have relatedly found that the links between ToM and EF did not extend to true-belief reasoning. Given this evidence, it is reasonable to suggest that EF might exclusively relate to recursive mental states involving a false-belief component, although some argue for a more general role for EF in ToM reasoning (e.g. Friedman & Leslie, 2004; Leslie et al., 2005; Leslie & Polizzi, 1998). In the studies presented here, we have only included tasks assessing the ability to ascribe beliefs about beliefs (whether true/false), and including other recursive mental states such as beliefs about desires or beliefs about emotions could shed light on whether the correlations we observed are specific to SOFB, or also generalize to other types of second-order mental states.

Methodological Alternatives

Much of the work conducted in the domain of adult first-order ToM used experimental rather than correlational methods to examine the role of EF in such ability (e.g., dual-tasks). This work, in comparison to the research reported here, generally observes that even simpler perspective-taking activities depend on a certain level of EF (e.g., Qureshi et al., 2010). As explained earlier, it may be that from a certain age, the role of EF must be examined through more subtle behaviour. Accordingly, the SOFB task introduced in this dissertation could be used in conjunction with an EF task, in order to examine the effect of taxing EF and ToM simultaneously. For example, Mckinnon and Moscovitch (2007) devised a dual-task paradigm in which a FOFB task, as well as a SOFB task, were to be completed under one of two conditions: A dual-task condition involving the presence of a concurrent auditory *N*-back task, and a control condition performed without the concurrent EF task. Especially in the SOFB condition, significantly more errors occurred under these dual-task requirements, suggesting that the EF task may have interfered with participants' SOFB reasoning. Building on these findings, dual-task effects might be found to be more or less pronounced depending on the nature of the EF task, which would provide a further indication of the specific components of EF most involved in SOFB (and/or FOFB). For instance, observing that concurrent working memory tasks disrupt SOFB performance above and beyond cognitive flexibility or inhibitory control tasks could provide additional support to the working memory explanation put forward in this dissertation.

Further, we did not find evidence for a link between inhibitory control and either FOFB or SOFB in this dissertation, and noted that this was somewhat unexpected given how robustly linked these constructs are in preschool. A nuanced interpretation of this finding may be that beyond preschool, inhibitory control manifests itself more implicitly than via the calculation of beliefs as such (Qureshi et al., 2020). Inhibitory failures are predictably encountered during online communication, for instance, when adults make egocentric errors in the way they assess another's knowledge. In typical common ground experiments, participants are instructed to move certain objects around a grid, according to the directives of a confederate who only sees some but not all of the objects in the grid. Adults' eye/reaching movements systematically evidence a tendency to consider one's privileged viewpoint first, by looking or reaching toward the object that only they can see. However, when more explicit measures are used (e.g., verbal predictions), almost no errors are detected (Begeer et al., 2010; Brown-Schmidt, 2009; Converse et al., 2008; Epley, 2008; Keysar et al., 2000; Keysar et al., 2003; Lin et al., 2010). Thus, perhaps what it takes for inhibitory control to significantly correlate with adults' false-belief reasoning are less explicit instructions to track another's beliefs, and more implicit focus on natural tendencies or subtle behaviour.

In related vein, one could manipulate the executive demands of the SOFB task (and/or FOFB) more directly, by increasing the salience of the object for example. It is well known that for children, the pull of reality is considerably stronger when they are aware of the target object's actual location, versus "low-inhibition" conditions in which the object has been taken away to an unknown location (e.g., Setoh et al., 2016; Southgate et al., 2007). In the version of the SOFB task administered here, the object of interest was housed in containers with an opaque front panel, such that participants could no longer see the object after the initial "hiding" procedure. The back of these containers were see-through, however, and perhaps on some trials, the scene could be reversed so that participants could see the object at all times, making it more difficult to inhibit. A design in which inhibitory control is directly manipulated could extend our findings beyond the correlational level and provide a more sensitive method to examine the influence of inhibitory control on ToM performance. Additionally, such manipulation could contribute to increasing the ecological validity of our task. Perhaps combined with additional modifications (e.g., providing a social context, including more human-like agents), presenting participants with more salient stimuli could be a proxy for the number of distractions typically present in real life interactions.

Beyond EF, What Other Factors Could Affect SOFB Performance?

Finally, while our results indicated that some of the variability in SOFB understanding may be linked to EF, the magnitude of the correlations found across studies remained relatively weak, suggesting that other factors are likely to be involved in explaining this variability. It is worth pointing out, however, that our samples consisted of highly educated university students, who are presumably at the higher end of both EF and ToM. The correlations observed here might thus be an underestimate of what might be found in community samples where both constructs may show more variability.

That aside, a number of studies have shown that social competence is an important correlate of both first and second-order ToM in children (Astington, 2003; Capage & Watson, 2001; Hughes & Leekam, 2004; Liddle & Nettle, 2006; Razza & Blair, 2009). For example, children who perform well on false-belief tasks tend to be better at coordinating with peers, communicate more successfully, and are more likely to experience positive social interactions (Astington, 2003). Similarly, there is also evidence showing that adults with larger social networks have higher mentalizing abilities (as assessed by the IMT) than adults with lower mentalizing abilities (Stiller & Dunbar, 2007). Beyond EF, then, there may be an important link between the extent of one's social competence and individual differences in SOFB. Perhaps adults who are more socially competent may find it easier to think more deeply about mental states, leading them to engage in recursive thinking more often, and perform better on SOFB tasks.

Another potential variable to consider for future studies is motivation. For instance, it is assumed that motivation affects people's inclination to engage in more systematic and elaborate cognitive processing (Güss et al., 2017; Thompson et al., 2000). Some findings indicate that manipulating participants' motivation - e.g., offering rewards based on performance - significantly increased the sophistication of reasoning in the domains of persuasion (Petty et al., 2015), and strategic negotiations (De Dreu et al., 2006). No study so far has examined the influence of motivation on SOFB reasoning, though it is relevant to note that recursive thinking tasks are often approached with some level of apprehended difficulty by participants. When asked to think through multiple iterations of other people's thoughts, participants often predict their own performance with pessimism, and report expecting the task to be difficult (de Weerd et al., 2017). Hence, motivation may play an important role in explaining why some individuals are more inclined to overcome this perceived difficulty and dedicate more effort to recursive mental state inferences. As mentioned previously, research on recursive thinking in strategic games evidences how only a small proportion of adults are willing to engage in a number of recursive steps when thinking about an opponent's likely actions (see Halevy, 2016 for a review). Under more explicit instructions to engage in recursive ToM, however, most adults do show an ability to consider five or more recursive mental states (as in the IMT for instance; e.g., Kinderman et al., 1998). Some motivational aspects may thus be important in explaining why individuals do have the capacity to think recursively about mental states, but vary in the extent to which they reliably use these skills across different contexts.

Conclusion

Taken together, our studies revealed important and specific links between EF and SOFB. We showed that the extent of the ToM-EF linkage found across the lifespan goes beyond the particular ability to attribute FOFB, to also include recursive forms of belief reasoning. Across three studies, the evidence we presented suggests positive relations between SOFB and EF, especially in the domain of working memory. We initially suggested that the mental state load of the SOFB task, as well as the recursive structure of the mental states involved may be potential drivers of these relations. When narrowed down to finer working memory components, our analysis revealed correlational patterns favouring a recursive hypothesis: The ability to manipulate information in working memory, as opposed to strictly storing this information, was most related to performance on the SOFB task. In our view, this finding suggests that the need to reason through the interaction of beliefs held by others, via a mental manipulation of the relevant information, may be the main factor driving the associations between working memory and SOFB. These suggestions are preliminary, but add to the ToM literature by documenting the ongoing involvement of EF in later-emerging and more complex forms of belief reasoning, as well as exploring, for the first time, whether different executive requirements exist between the ability to attribute beliefs, and the ability to attribute recursive beliefs in adulthood. Studies of adults are important in providing an account of the extent to which EF is involved in the maintenance of mature ToM (Apperly et al., 2009), and we hope that the current work provides a motivation for future research in this direction.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Abrams, D., Rutland, A., Palmer, S. B., Pelletier, J., Ferrell, J., & Lee, S. (2014). The role of cognitive abilities in children's inferences about social atypicality and peer exclusion and inclusion in intergroup contexts. *British Journal of Developmental Psychology, 32*(3), 233-247.

Allen, T. A., Rueter, A. R., Abram, S. V., Brown, J. S., Deyoung, C. G., & Mõttus, R. (2017). Personality and neural correlates of mentalizing ability. *European journal of personality, 31*(6), 599-613.

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

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology, 65*(5), 825-839.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006a). Is belief reasoning automatic?. *Psychological Science, 17*(10), 841-844.

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2006b). Intact first-and second order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience*, *1*(3-4), 334-348.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental psychology, 45*(1), 190.

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. *Child development, 82*(5), 1691-1703.

Archibald, L. M., Levee, T., & Olino, T. (2015). Attention allocation: Relationships to general working memory or specific language processing. *Journal of Experimental Child Psychology, 139*, 83- 98.

*Arslan, B., Hohenberger, A., & Verbrugge, R. (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PloS one, 12*(1), e0169510.

Arslan, B., Taatgen, N. A., & Verbrugge, R. (2017). Five-year-olds' systematic errors in second order false belief tasks are due to first-order theory of mind strategy selection: a computational modeling study. *Frontiers in psychology, 8,* 275.

Astington, J.W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi, V. Slaughter (Eds.), *Individual differences in theory of mind*. (pp.12-38). Macquarie monographs in cognitive science, Psychology Press.

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New ideas in Psychology, 20*(2-3), 131-144.

Austin, G., Groppe, K., & Elsner, B. (2014). The reciprocal relationship between executive function and theory of mind in middle childhood: A 1-year longitudinal perspective. *Frontiers in psychology, 5*, 655.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54-70.

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

Badre, D., & Wagner, A. D. (2006). Computational and neurobiological mechanisms underlying cognitive flexibility. *Proceedings of the National Academy of Sciences, 103*(18), 7186 7191.

Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences, 1118*(1), 122-141.

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2019). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*.

Band, G. P., van der Molen, M. W., Overtoom, C. C., & Verbaten, M. N. (2000). The ability to activate and inhibit speeded responses: Separate developmental trends. *Journal of experimental child psychology, 75*(4), 263-290.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child development, 82*(6), 1887 1905.

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., & Brass, M. (2017). Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Social cognitive and affective neuroscience, 12*(3), 391-400.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child psychology and Psychiatry, 38*(7), 813-822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind". *Cognition, 21*(1), 37-46.

Barrett, H. C., Cosmides, L., & Tooby, J. (2010). Coevolution of cooperation, causal cognition and mindreading. *Communicative & Integrative Biology, 3*(6), 522-524.

Bates B.T., Dufek J.S., Davis H.P. (1992). The effect of trial size on statistical power. Medicine and *Science in Sports and Exercise, 24*, 1059-1065.

Baumeister, R. F., & Vohs, K. D. (2003). Self-regulation and the executive function of the self. *Handbook of self and identity, 1*, 197-217.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71.

Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using theory of mind to represent and take part in social interactions: Comparing individuals with high functioning autism and typically developing controls. *European Journal of Developmental Psychology, 7*(1), 104-122.

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research, 37*(5), 481-502.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development, 81*(6), 1641-1660.

*Bibby, H., & McDonald, S. (2005). Theory of mind after traumatic brain injury. *Neuropsychologia,43*(1), 99-114.

Birch, S. A., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in cognitive sciences, 8*(6), 255-260.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382-386.

Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of child psychology and psychiatry, 47*(3-4),296-312.

*Bock, A. M., Gallaway, K. C., & Hund, A. M. (2015). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development, 16*(3), 509-521.

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning disabilities,43*(6), 541-552.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis, 2*, 221-235.

Borowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia,5*, 135–140.

*Bozikas, V. P., Giannakou, M., Kosmidis, M. H., Kargopoulos, P., Kioseoglou, G., Liolios, D., & Garyfallos, G. (2011). Insights into theory of mind in schizophrenia: the impact of cognitive impairment. *Schizophrenia research, 130*(1-3), 130-136.

Bradford, E. E., Jentzsch, I., & Gomez, J. C. (2015). From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition, 138*, 21-34.

Brass, M., & Cramon, D. Y. V. (2004). Decomposing components of task preparation with functional magnetic resonance imaging. *Journal of cognitive neuroscience, 16*(4), 609 620.

Braüner, T., Blackburn, P., & Polyanskaya, I. (2016, August). Second-order false-belief tasks: Analysis and formalization. In *International Workshop on Logic, Language, Information, and Computation* (pp. 125-144). Springer, Berlin, Heidelberg.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5*(1), 49-62.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental neuropsychology, 26*(2), 571-593.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., ... & Lynam, D. R. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. *Developmental psychology, 39*(2), 222.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic bulletin & review, 16*(5), 893-900.

*Brunet, M. (2013). *Deception in children with and without severe conduct problems* [Doctoral dissertation, University of Toronto]. ProQuest Dissertations and Theses Global.

Burchardi, K. B., & Penczynski, S. P. (2014). Out of your mind: Eliciting individual reasoning in one shot games. *Games and Economic Behavior, 84*, 39-57.

Buttelmann, F., & Karbach, J. (2017). Development and plasticity of cognitive flexibility in early and middle childhood. *Frontiers in psychology, 8*, 1040.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*(5), 606-637.

Byrne, R. & Whiten, A., eds. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Clarendon.

*Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: Irony understanding and recursive theory of mind. *Research in developmental disabilities, 35*(11),3191-3198.

*Caillies, S., Hody, A., & Calmus, A. (2012). Theory of mind and irony comprehension in children with cerebral palsy. *Research in developmental disabilities, 33*(5), 1380-1388.

Caillies, S., & Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. *Developmental Science, 11*(5), 703-711.

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. *Trends in cognitive sciences, 7*(5), 225-231.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2015). A psychological approach to strategic thinking in games. *Current Opinion in Behavioral Sciences, 3*, 157-162.

Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(4), 591.

Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education and Development, 12*(4), 613-628.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186-197.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental psychology, 40*(6), 1105.

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

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 73-92.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child development, 69*(3), 672 691.

Casey, B. J., Galvan, A., & Hare, T. A. (2005). Changes in cerebral functional organization during cognitive development. *Current opinion in neurobiology, 15*(2), 239-244.

*Cassetta, B. D., Pexman, P. M., & Goghari, V. M. (2018). Cognitive and Affective Theory of Mind and Relations With Executive Functioning in Middle Childhood. *Merrill Palmer Quarterly, 64*(4), 514-538.

Čavojová, V., Belovičová, Z., & Sirota, M. (2011). Mindreading and empathy as predictors of prosocial behavior. *Studia psychologica, 53*(4), 351-362.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology,23*(2), 201-216.

Chelonis, J., Daniels-Shawb, J., Blakea, D., & Paule, M. (2000). Developmental aspects of delayed matching-to-sample task performance in children. *Neurotoxicology and Teratology, 22*, 683-694.

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics: The evolution of a social mind*. University of Chicago Press.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods, 19*(2), 211.

Chomsky, N. (1957). Logical structures in language. *American Documentation (pre-1986), 8*(4), 284.

Chomsky, N. (1995). Language and nature. *Mind, 104*(413), 1-61.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Cohen, A. S., & German, T. C. (2009). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition, 115*(3), 417-425.

Cohen, J. (1988). The effect size. *Statistical power analysis for the behavioral sciences*, 77-83.

Colé, P., Duncan, L. G., & Blaye, A. (2014). Cognitive flexibility predicts early reading skills. *Frontiers in Psychology, 5*, 565.

*Comay, J. (2009). *Individual differences in narrative perspective-taking and theory of mind* [Doctoral Dissertation, University of Toronto]. ProQuest Dissertations and Theses Global.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology, 31*, 103 128.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: mood affects theory-of-mind use. *Emotion, 8*(5), 725.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review, 12*(5), 769-786.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7*(12), 547-552.

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in psychology, 8*, 1482.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing* (pp. 75-176). Academic Press.

Corballis, M. C. (2014). *The recursive mind: The origins of human language, thought, and civilization updated edition*. Princeton University Press.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34* (02), 891B. (University Microfilms No. AAI05–77717).

Crane, P. K., Narasimhalu, K., Gibbons, L. E., Mungas, D. M., Haneuse, S., Larson, E. B., ... & van Belle, G. (2008). Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of clinical epidemiology, 61*(10), 1018-1027.

Crescioni, A., Ehrlinger, J., Alquist, J. L., Conlon, K. E., Baumeister, R. F., Schatschneider, C., & Dutton, G. R. (2011). High trait self-control predicts positive health behaviors and success in weight loss. *Journal of health psychology, 16*(5), 750-759.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications forclinical and developmental neuroscience. *Trends in neurosciences, 38*(9), 571-578.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language, 19*(4), 450.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078.

Davis, M. (1965). *The Undecidable. Basic papers on undecidable propositions, unsolvable problems, and computable functions*. Raven Press.

De Dreu, C. K., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of personality and socialpsychology, 90*(6), 927.

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the CANTAB. I: development of executive function over the lifespan. *Journal of clinical and experimental neuropsychology, 25*(2), 242-254.

Dennett, D. (1978). Beliefs about beliefs. *Behavior and Brain Sciences*, *4*, 568–70.

Denson, T. F., Pedersen, W. C., Friese, M., Hahm, A., & Roberts, L. (2011). Understanding impulsive aggression: Angry rumination and reduced self-control capacity are mechanisms underlying the provocation-aggression relationship. *Personality and Social Psychology Bulletin, 37*(6), 850-862.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain and cognition, 41*(1), 66-86.

Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of mind: did evolution fool us?. *PloS One, 9*(2), e87619.

Devaine, M., San-Galli, A., Trapanese, C., Bardino, G., Hano, C., Saint Jalme, M., ... & Daunizeau, J. (2017). Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species. *PLoS computational biology, 13*(11), e1005833.

De Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858 1878.

De Villiers, J. G., Hobbs, K., & Hollebrandse, B. (2014). Recursive complements and propositional attitudes. In *Recursion: Complexity in cognition* (pp. 221-242). Springer, Cham.

De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive development, 17*(1), 1037-1060.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child development, 85*(5), 1777-1794.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*(5), 758.

de Weerd, H., Verbrugge, R., & Verheij, B. (2013, December). Higher-order theory of mind in negotiations under incomplete information. In *International Conference on Principles and Practice of Multi-Agent Systems* (pp. 101-116). Springer, Berlin, Heidelberg.

de Weerd, H., Verbrugge, R., & Verheij, B. (2017). Negotiating with other minds: the role of recursive theory of mind in negotiation with incomplete information. *Autonomous Agents and Multi Agent Systems, 31*(2), 250-287.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of frontal lobe function*, 466 503.

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168.

Dick, A. S. (2014). The development of cognitive flexibility beyond the preschool period: an investigation using a modified Flexible Item Selection Task. *Journal of experimental child psychology, 125*, 13-34.

Di Tella, M., Ardito, R. B., Dutto, F., & Adenzato, M. (2020). On the (lack of) association between theory of mind and executive functions: a study in a non-clinical adult sample. *Scientific Reports, 10*(1), 1-9.

Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., & Sommer, M. (2012). Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *Neuroimage, 60*(3), 1652-1661.

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science, 11*(1), 133-155.

Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of mind and executive function in Chinese preschool children. *Developmental psychology, 52*(4), 582.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental science, 13*(2), 331-338.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 6*(5), 178-190.

Dunbar, R. I. (2009). The social brain hypothesis and its implications for social evolution.*Annals of human biology, 36*(5), 562-572.

Dunbar, R. I. (2011). Evolutionary basis of the social brain. *Oxford handbook of social neuroscience*, 28-38.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology, 43*(6), 1428.

*Ebert, S. (2020). Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of Experimental Child Psychology, 191*, 104739.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj, 315*(7109), 629-634.

Eliot, J., Lovell, K., Dayton, C. M., & McGrady, B. F. (1979). A further investigation of children's understanding of recursive thinking. *Journal of Experimental Child Psychology, 28*(1), 149- 157.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General, 128*(3), 309.

Epley, N. (2008). Solving the (real) other minds problem. *Social and personality psychology compass, 2*(3), 1455-1474.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*(6), 760-768.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology, 146*(4), 348-361.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience, 14*(3), 340 347.

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2014). Cognitive flexibility, theory of mind, and hyperactivity/inattention. *Child development research*, *2014*.

Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and individual differences, 49*(5), 414-418.

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology, 70*(8), 1646-1660.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child development, 79*(1), 126-138.

Fizke, E., Barthel, D., Peters, T., & Rakoczy, H. (2014). Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition*, *130*(3), 315-334.

Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information, 17*(4), 417-442.

*Ford, R. M., Griffiths, S., Neulinger, K., Andrews, G., Shum, D. H., & Gray, P. H. (2017). Impaired prospective memory but intact episodic memory in intellectually average 7-to 9-year-olds born very preterm and/or very low birth weight. *Child Neuropsychology, 23*(8), 954-979.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: General, 133*(1),101.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186-204.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological science, 17*(2), 172 179.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of experimental psychology: General, 137*(2), 201.

Friedman, O., & Leslie, A. M. (2004). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science, 15*(8), 547-552.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief-desire reasoning: Inhibition or general difficulty?. *Developmental Science, 8*(3), 218-225.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale.

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

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review, 17*(5), 673-679.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin, 134*(1), 31.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. Developmental Psychology, 40, 177-190.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*(1), 129-152.

Germine, L., Dunn, E. C., McLaughlin, K. A., & Smoller, J. W. (2015). Childhood adversity is associated with adult theory of mind and social affiliation, but not face processing. *PloS one, 10*(6), e0129612.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Rapoport, J. L. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, 101*(21), 8174-8179.

Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making, 25*(1), 95-108.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In Tooby, J., Leslie, A.M., Sperber, D., Caramazza, A., Hills, A.E., Leek, E.C., …& Cosmides, L. (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). Cambridge University Press.

Guajardo, N. R., & Cartwright, K. B. (2016). The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers' language comprehension and later readingawareness and comprehension in elementary school. *Journal of Experimental Child Psychology, 144*, 27-45.

Güss, C. D., Burger, M. L., & Dörner, D. (2017). The role of motivation in complex problem solving. *Frontiers in psychology, 8*, 851.

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development,4*(3), 275-298.

Halevy, N. (2016). Strategic thinking. In *Advances in Experimental Social Psychology* (Vol. 54, pp. 1-66). Academic Press.

Happaney, K., Zelazo, P. D., & Stuss, D. T. (2004). Development of orbitofrontal function: Current themes and future directions. *Brain and cognition, 55*(1), 1-10.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101-119.

Harnishfeger, K. K., & Bjorklund, D. F. (1994). A developmental perspective on individual differences in inhibition. *Learning and individual differences, 6*(3), 331-355.

Hebb, D. O. (1961). Brain mechanisms and learning. In J. F. Delafresnaye (Ed.), *Distinctive features of learning in the higher animal* (pp. 37-46). Oxford University Press.

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition, 85*(1), 1-36.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. *Research synthesis methods, 1*(1), 39-65.

*Henry, A., Tourbah, A., Chaunu, M. P., Bakchine, S., & Montreuil, M. (2017). Social cognition abilities in patients with different multiple sclerosis subtypes. *Journal of the International Neuropsychological Society, 23*(8), 653-664.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj, 327*(7414), 557-560.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*, 184–198.

Holden, J., Francisco, E., Lensch, R., Tommerdahl, A., Kirsch, B., Zai, L., ... & Tommerdahl, M. (2019). Accuracy of different modalities of reaction time testing: Implications for online cognitive assessment tools. *bioRxiv*, 726364.

Hollebrandse, B., Hobbs, K., & De Villiers, J., Roeper, T. (2008). Second order embedding and second order false belief. In *Language Acquisition and Development, Proceedings of GALA 2007*.

*Hsu, Y. K., & Cheung, H. (2013). Two mentalizing capacities and the understanding of two types of lie telling in children. *Developmental Psychology*, *49*(9), 1650.

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. (2002). Executive functions and development: Emerging themes. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 201-209.

Hughes, C., & Cutting, A. L. (1999). Nature, nurture, and individual differences in early understanding of mind. *Psychological Science, 10*(5), 429-432.

Hughes, C., & Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives, 9*(3), 149 153.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair?. *Developmental neuropsychology, 28*(2), 645-668.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental psychology, 43*(6), 1447.

Hughes, C., Ensor, R., Wilson, A., Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20 36.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions?. *Child and adolescent mental health, 7*(3), 131-142.

Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture?. *Child development, 76*(2), 356-370.

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social development, 13*(4), 590-619.

Huizinga, M., & Van der Molen, M. W. (2007). Age-group differences in set-switching and set maintenance on the Wisconsin Card Sorting Task. *Developmental neuropsychology, 31*(2), 193-215.

*Hur, J. W., Byun, M. S., Shin, N. Y., Shin, Y. S., Kim, S. N., Jang, J. H., & Kwon, J. S. (2013). General intellectual functioning as a buffer against theory-of-mind deficits in individuals at ultra-high risk for psychosis. *Schizophrenia research, 149*(1-3), 83-87.

*Huyder, V., & Nilsen, E. S. (2012). A dyadic data analysis of executive functioning and children's socially competent behaviours. *Journal of applied developmental psychology, 33*(4), 197- 208.

Ibanez, A., Huepe, D., Gempp, R., Gutiérrez, V., Rivera-Rei, A., & Toledo, M. I. (2013). Empathy, sex and fluid intelligence as predictors of theory of mind. *Personality and Individual Differences, 54*(5), 616-621.

Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before?. *Journal of Experimental Child Psychology, 149*, 98-115.

*Kazi, S., Kazali, E., Makris, N., Spanoudis, G., & Demetriou, A. (2019). Cognizance in cognitive development: A longitudinal study. *Cognitive Development, 52*, 100805.

Kemps, E., De Rammelaere, S., & Desmet, T. (2000). The development of working memory: Exploring the complementarity of two models. *Journal of Experimental Child Psychology, 77*, 89–109.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*, 72–88.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

*Kim, Y. S. G. (2020). Theory of mind mediates the relations of language and domain-general cognitions to discourse comprehension. *Journal of Experimental Child Psychology, 194*, 104813.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191-204.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta- analysis. *Research Synthesis Methods, 2*(1), 61-76.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus response compatibility--a model and taxonomy. *Psychological review, 97*(2), 253.

*Kretschmer, A., Lampmann, S. A., & Altgassen, M. (2014). Relations between moral reasoning, theory of mind and executive functions in children with autism spectrum disorders. International *Journal of Developmental Disabilities, 60*(3), 174-183.

Lagattuta, K. H., Kramer, H. J., Kennedy, K., Hjortsvang, K., Goldfarb, D., & Tashjian, S. (2015). Beyond Sally's missing marble: Further development in children's understanding of mind and emotion in middle childhood. In *Advances in child development and behavior* (Vol. 48, pp. 185-217). JAI.

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six-to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology, 46*(6), 1417.

Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*(2), 659-674.

Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., & Farah, M. J. (2018). Childhood socioeconomicstatus and executive function in childhood and beyond. *PloS one, 13*(8), e0202964.

Launay, J., Pearce, E., Wlodarski, R., van Duijn, M., Carney, J., & Dunbar, R. I. (2015). Higher order mentalising and executive functioning. *Personality and individual differences, 86*, 6-14.

*Lavoie, J., & Talwar, V. (2020). Care to share? Children's cognitive skills and concealing responses to a parent. *Topics in Cognitive Science, 12*(2), 485-503.

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology, 126*, 52-67.

Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology*, *163*, 69-86.

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

Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience, 114*(10), 1291-1311.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. *Mapping the mind: Domain specificity in cognition and culture, 29*, 119-48.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in cognitive sciences, 8*(12), 528-533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive psychology, 50*(1), 45-85.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental science, 1*(2), 247-253.

Lewis, D. (1996). *Convention*. Harvard University Press.

Lewis, P. A., Birch, A., Hall, A., & Dunbar, R. I. (2017). Higher order intentionality tasks are cognitively more demanding. *Social cognitive and affective neuroscience, 12*(7), 1063 1071.

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage, 57*(4), 1624-1629.

Li, H., & Leung, M. T. (2020). Relations between verb factivity and first-order and second-order false belief understanding: Evidence from Mandarin-speaking typically developing children and children with autism spectrum disorders. *Clinical Linguistics & Phonetics, 34*(1-2), 185-200.

*Li, X., Wang, K., Wu, J., Hong, Y., Zhao, J., Feng, X., ... & Zhang, X. (2014). The link between impaired theory of mind and executive function in children with cerebral palsy. *Research in Developmental Disabilities, 35*(7), 1686-1693.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school age children. *Journal of Cultural and Evolutionary Psychology, 4*(3-4), 231-244.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46*(3), 551-556.

Lockl, K., & Schneider, W. (2007). Knowledge about the mind: Links between theory of mind and later metamemory. *Child Development, 78*(1), 148-167.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological review, 91*(3), 295.

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition, 170*, 25-30.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*(3), 184-189.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*(1), 1-13.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic bulletin & review, 2*(2), 174-207.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child development, 76*(3), 697-712.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4-to 12-year-old children. *Developmental neuropsychology, 22*(3), 595-624.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child development, 75*(5), 1357 1372.

Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., ... & Sweeney, J. A.(2001). Maturation of widely distributed brain function subserves cognitive development. *Neuroimage, 13*(5), 786-793.

Lyons-Warren, A., Lillie, R., & Hershey, T. (2004). Short-and long-term spatial delayed response performance across the lifespan. *Developmental neuropsychology, 26*(3), 661 678.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin, 109*(2), 163.

Magnus, B. E., Willoughby, M. T., Blair, C. B., & Kuhn, L. J. (2019). Integrating item accuracy and reaction time to improve the measurement of inhibitory control abilities in early childhood. *Assessment, 26*(7), 1296-1306.

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development, 33*, 40–55.

Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false belief in children: What changes when mothers rather than siblings speak?. *European Journal of Developmental Psychology, 10*(3), 301-317.

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1124 –1140.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual review of psychology, 59*.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102*(2), 179-218.

Meijering, B., Van Maanen, L., Van Rijn, H., & Verbrugge, R. (2010). The facilitative effect of context on second-order social reasoning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Meijering, B., Van Rijn, H., Taatgen, N., & Verbrugge, R. (2011). I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Meijering, B., Van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2013). Reasoning about diamonds, gravity and mental states: The cognitive costs of theory of mind. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Meltzoff, A. N. (2011). Social cognition and the origins of imitation, empathy, and theory of mind. *The Wiley-Blackwell handbook of childhood cognitive development, 1*, 49-75.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology, 90*(2), 227.

Miller, H. V., Barnes, J. C., & Beaver, K. M. (2011). Self-control and health outcomes in a nationally representative sample. *American journal of health behavior, 35*(1), 15-27.

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.

Miller, P. H., Kessel, F. S., & Flavell, J. H. (1970). Thinking about people thinking about people thinking about...: A study of social cognitive development. *Child Development*, 613-623.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the structure of behavior*. Holt, Rinehart & Winston.

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological bulletin, 135*(5), 749.

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. Psychology Press.

Miller, S. E., Reavis, R. E., & Avila, B. N. (2018). Associations Between Theory of Mind, Executive Function, and Friendship Quality in Middle Childhood. *Merrill-Palmer Quarterly, 64*(3), 397-426.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child development, 78*(2), 622-646.

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000a). Assessment of executive functions in clinicalsettings: Problems and recommendations. In *Seminars in speech and language* (Vol. 21, No. 02, pp. 0169-0183). Copyright© 2000 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4663.

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

Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., Sears, M.R., Thomson, W.M., and Caspi, A. (2011). A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7) 2693-2698.

Moher D, Liberati A, Tetzlaff J, Altman DG. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med 6*(7): e1000097.

Moses, L.J. (2001). Executive accounts of theory-of-mind development. *Child development*, *72*(3), 688-690.

Moses, L. J. (2005). Executive functioning and children’s theories of mind. *Other minds: How humans bridge the divide between self and others*, 11-25.

Moses, L. J., Carlson, S. M., Stieglitz, S., & Claxton, L. J. (2003). Executive function, prepotency, and children’s theory of mind. *Unpublished manuscript*.

Moses, L. J., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theories of mind. *Self and social regulation: Social interaction and the development of social understanding and executive functions*, 218-233.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of neuroscience methods, 222*, 250-259.

Müeller, U., & Overton, W. F. (2010). Thinking about thinking—Thinking about measurement: A Rasch analysis of recursive thinking. *Journal of Applied Measurement, 11*(1), 78.

Müller, U., Baker, L., & Yeung, E. (2013). A developmental systems approach to executive function. *Advances in child development and behavior, 45*, 39-66.

Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of experimental child psychology, 111*(2), 331-348.

Mutter, B., Alcorn, M. B., & Welsh, M. (2006). Theory of mind and executive function: Working memory capacity and inhibitory control as predictors of false-belief task performance. *Perceptual and Motor skills, 102*(3), 819-835.

Myers, L., & Sirois, M. J. (2006). Spearman Correlation Coefficients. Differences between. *Encyclopedia of Statistical Sciences*, *12*.

Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*(5), 1253-1274.

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social perceptual, theory of mind. *European Journal of Personality: Published for the European Association of Personality Psychology, 22*(4), 323-335.

Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016). Measuring mentalizing ability: A within subject comparison between an explicit and implicit version of a ball detection task. *PloS one,11*(10), e0164373.

Nilsen, E. S., & Bacso, S. A. (2017). Cognitive and behavioural predictors of adolescents' communicative perspective-taking and social relationships. *Journal of Adolescence, 56*, 52-63.

*Nilsen, E. S., & Valcke, A. (2018). Children's sharing with collaborators versus competitors: The impact of theory of mind and executive functioning. *Journal of Applied Developmental Psychology, 58*, 38-48.

Oberauer, K., Süβ, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence?. *Intelligence, 36*(6), 641-652.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics, 43*, 95-106.

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior, 36*(4), 313- 322.

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development, 79*(1), 80-99.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *science, 308*(5719), 255-258.

Oppenheimer, L. (1986). Development of recursive thinking: Procedural variations. *International Journal of Behavioral Development, 9*(3), 401-411.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of educational statistics, 8*(2), 157-159.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews, 5*(1), 210.

*Paine, A. L., Pearce, H., van Goozen, S. H., de Sonneville, L. M., & Hay, D. F. (2018). Late, but not early, arriving younger siblings foster firstborns' understanding of second-order false belief. *Journal of Experimental Child Psychology, 166*, 251-265.

Partington, J. E., & Leiter, R. G. (1949). Partington's Pathways Test. *Psychological Service Center Journal*.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental psychology, 46*(2), 530.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). Unified tri-services cognitive performance assessment battery: review and methodology (AAMRL-TR 87–007). *Dayton: Armstrong Aerospace Medical Research Laboratory*.

Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. *Developing theories of mind*, 271-294.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 270- 283). Cambridge University Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in cognitive sciences, 3*(9), 337-344.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child development, 73*(3), 752-767.

*Perner, J., Kain, W., & Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 141-158.

Perner, J., Stummer, S., & Lang, B. (1999). Executive functions and theory of mind: Cognitive complexity or functional dependence?. In *Developing theories of intention: Social understanding and self-control* (pp. 133-152). Psychology Press.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second order beliefs by 5-to 10-year-old children. *Journal of experimental child psychology*, *39*(3), 437- 471.

Perugini, E. M., Harvey, E. A., Lovejoy, D. W., Sandstrom, K., & Webb, A. H. (2000). The predictive power of combined neuropsychological measures for attention deficit/hyperactivity disorder in children. *Child Neuropsychology, 6*(2), 101-114.

Pesch, A., Semenov, A. D., & Carlson, S. M. (2020). The Path to Fully Representational Theory of Mind: Conceptual, Executive, and Pragmatic Challenges. *Frontiers in Psychology, 11*, 3020.

*Peskin, J., Prusky, C., & Comay, J. (2014). Keeping the reader's mind in mind: development of perspective-taking in children's dictations. *Journal of applied developmental psychology, 35*(1), 35-43.

Petty, R. E., Cacioppo, J. T., & Kasmer, J. A. (2015). The role of affect in the elaboration likelihood model of persuasion. In *Communication, Social Cognition, and Affect (PLE: Emotion)* (pp. 133-162). Psychology Press.

Phillips, L. H., Bull, R., Allen, R., Insch, P., Burr, K., & Ogg, W. (2011). Lifespan aging and belief reasoning: Influences of executive function and social cue decoding. *Cognition, 120*(2), 236- 247.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it?. *Cognition, 95*(2), 201-236.

Polyanskaya, I., Braüner, T., & Blackburn, P. R. (2018). Second-Order False Beliefs and Recursive Complements in Children with ASD. In *42nd annual Boston University Conference on Language Development* (pp. 632-643). Cascadilla Press.

Powell, J. L., Lewis, P. A., Dunbar, R. I., García-Fiñana, M., & Roberts, N. (2010). Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia, 48*(12), 3554-3562.

Powell, J. L., Kemp, G. J., Dunbar, R. I., Roberts, N., Sluming, V., & García-Fiñana, M. (2014). Different association between intentionality competence and prefrontal volume in left and right handers. *Cortex, 54*, 63-76.

Powell, J., Lewis, P. A., Roberts, N., Garcia-Finana, M., & Dunbar, R. I. (2012). Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2157-2162.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences, 1*(4), 515-526.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual task study of adults. *Cognition, 117*(2), 230-236.

Qureshi, A. W., Monk, R. L., Samson, D., & Apperly, I. A. (2020). Does interference between self and other perspectives in theory of mind tasks reflect a common underlying process?

Evidence from individual differences in theory of mind and inhibitory control. *Psychonomic bulletin & review, 27*(1), 178-190.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

*Raijmakers, M. E., Mandell, D. J., van Es, S. E., & Counihan, M. (2014). Children's strategy use when playing strategic games. *Synthese, 191*(3), 355-370.

Rapport, L. J., Friedman, S. L., Tzelepis, A., & Van Voorhis, A. (2002). Experienced emotion and affect recognition in adult attention-deficit hyperactivity disorder. *Neuropsychology, 16*(1), 102.

Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*(3), 332-343.

Riggs, N. R., Spruijt-Metz, D., Sakuma, K. L., Chou, C. P., & Pentz, M. A. (2010). Executive cognitive function and food intake in children*. Journal of nutrition education and behavior, 42*(6), 398- 403.

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. E. (1998). *The prefrontal cortex: executive and cognitive functions*. Oxford University Press.

Ronald, A., Viding, E., Happé, F., & Plomin, R. (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: A twin study. *Social Neuroscience, 1*(3-4), 412-425.

*Rostad, K., & Pexman, P. M. (2014). Developing appreciation for ambivalence: The understanding of concurrent conflicting desires in 4-to 7-year-old children. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale, 68*(2), 122.

Rothbart, M. K., & Posner, M. I. (1985). Temperament and the development of self-regulation. In *The neuropsychology of individual differences* (pp. 93-123). Springer.

Rothmayr, C., Sodian, B., Hajak, G., Döhnel, K., Meinhardt, J., & Sommer, M. (2011). Common and distinct neural networks for false-belief reasoning and inhibitory control. *NeuroImage, 56*(3), 1705-1713.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental review, 34*(3), 265-293.

Rushworth, M. F., Passingham, R. E., & Nobre, A. C. (2005). Components of attentional set switching. *Experimental psychology, 52*(2), 83-98.

Russell, J. (1996). *Agency: Its role in mental development*. Taylor & Francis.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British journal of developmental psychology, 9*(2), 331-349.

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General, 144*(5), 898.

Saarni, C. (1999). *The development of emotional competence*. Guilford press.

*Saban-Bezalel, R., Dolfin, D., Laor, N., & Mashal, N. (2019). Irony comprehension and mentalizing ability in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders, 58*, 30-38.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and US preschoolers. *Psychological science, 17*(1), 74-81.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal ofExperimental Psychology: Human Perception and Performance, 36*(5), 1255.

Samuel, S., Roehr-Brackin, K., Jelbert, S., & Clayton, N. S. (2019). Flexible egocentricity: Asymmetric switch costs on a perspective-taking task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(2), 213.

Santos, L. R., Flombaum, J. I., & Phillips, W. (2007). The evolution of human mindreading: How non- human primates can inform social cognitive neuroscience. *Evolutionary cognitive neuroscience*, 433-456.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press.

Schneider, W., Lockl, K., & Fernandez, O. (2005). Interrelationships among theory of mind, executive control, language development, and working memory in young children: A longitudinal analysis. Young children's cognitive development: *Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*, 259-284.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237-249.

Scott-Phillips, T.C. (2015). *Speaking our minds*. Palgrave Macmillan.

Searle, J. R. (1979). What is an intentional state?. *Mind, 88*(349), 74-92.

Searle, J. R. (1995). *The construction of social reality*. Free Press.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences, 113*(47), 13360-13365.

Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents' emotion regulation in daily life: Links to depressive symptoms and problem behavior. *Child development, 74*(6), 1869 1880.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81*, 174–176.

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus response correspondence and age. *Ergonomics, 6*(1), 99-105.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General, 143*(2), 534.

Smith, E.E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657-1661.

Smith, M., Apperly, I., & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Development, 74*(6), 1709-1719.

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Frontiers in psychology, 6*, 328.

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development, 84*(6), 1846-1854.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological science, 18*(7), 587-592.

Spencer-Smith, M., & Anderson, V. (2009). Healthy and abnormal development of the prefrontal cortex. *Developmental Neurorehabilitation, 12*(5), 279-297.

Sperber, D. (1994). The epidemiology of representations. *Mapping the mind: Domain specificity in cognition and culture*, 39.

Sperber, D. (2000). *Metarepresentations in an evolutionary perspective*. Oxford University Press.

Sternberg, S. (2010). Reaction-time experimentation. *Psychology, 600*, 301.

Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks, 29*(1), 93-104.

*Strasser, K., & Río, F. D. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly, 49*(2), 169-187.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Stuss, D.T., & Benson, D.F. (1986). *The Frontal Lobes*. Raven Press.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261-288.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second order beliefs. *Developmental Psychology*, *30*, 395–402.

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child development, 83*(2), 452-460.

Tabachnick, B.G., Fidell, L.S. (1996). *Using multivariate statistics.* Harper Collins.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481-498). Pearson.

Thompson, E. P., Kruglanski, A. W., & Spiegel, S. (2000). Attitudes as knowledge structures and persuasion as a specific case of subjective knowledge acquisition. *Why we evaluate: Functions of attitudes, 466*.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta regression. *Psychological methods, 20*(3), 375.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological bulletin, 133*(6), 1038.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635-654.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive psychology, 71*, 1-26.

*Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of mind development in adolescence and early adulthood: the growing complexity of recursive thinking ability. *Europe's journal of psychology, 11*(1), 112.

van den Bos, E., de Rooij, M., Sumter, S. R., & Westenberg, P. M. (2016). Continued development of recursive thinking in adolescence: Longitudinal analyses with a revised recursive thinking test. *Cognitive Development*, *37*, 28-41.

van Mourik, R., Oosterlaan, J., & Sergeant, J. A. (2005). The Stroop revisited: A meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry, 46*(2), 150-165.

Verbrugge, R. (2009). Logic and social cognition. *Journal of Philosophical Logic, 38*(6), 649 680.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in cognitive sciences, 12*(11), 418-424.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software, 36*(3), 1-48.

*Vilenskaya, G. A., & Lebedeva, E. I. (2017). The Relations Between Theory of Mind and Behavioral Control Among Primary School Age Children. *The Russian Journal of Cognitive Science, 4*(4), 29-36.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review, 20*(4), 766-772.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., ... & Weersing, V. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist, 72*(2), 79.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development, 72*(3), 655-684.

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and of thoughts. *Child Development, 67*(3), 768-788.

Welsh, M., & Schmitt-Wilson, S. (2013). Executive function, identity, and career decision making in college students. *SAGE Open, 3*(4), 2158244013505755.

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.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental psychology, 35*(1), 205.

*Williams, S., Moore, K., Crossman, A. M., & Talwar, V. (2016). The role of executive functions and theory of mind in children's prosocial lie-telling. *Journal of Experimental Child Psychology, 141*, 256-266.

Wimmer, H. (1989). Common-sense Mentalismus und Emotion: einige entwicklungspsychologische Implikationen. In *Denken und Fühlen* (pp. 56-66). Springer.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Winstok, Z. (2009). From self-control capabilities and the need to control others to proactive and reactive aggression among adolescents. *Journal of Adolescence, 32*(3), 455-466.

Woodruff, G., & Premack, D. (1979). Intentional communication in the chimpanzee: The development of deception. *Cognition, 7*(4), 333-362.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*(4), 954-959.

Zabelina, D. L., Friedman, N. P., & Andrews-Hanna, J. (2019). Unity and diversity of executive functions in creativity. *Consciousness and cognition, 68*, 47-56.

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

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). Executive Function: Implications for Education. NCER 2017-2000. *National Center for Education Research*.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. *In Handbook of Developmental Cognitive Neuroscience*. MIT Press.

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child development perspectives, 6*(4), 354-360.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of general psychology, 1*(2), 198-226.

Zelazo, P. D., & Müller, U. (2002). The balance beam in the balance: Reflections on rules, relational complexity, and developmental processes. *Journal of Experimental Child Psychology, 81*(4), 458-465.

Appendix A

Table S1

Studies Included in Meta-Analysis (Study 1)

Study	Country	Age (years)	*N*	SOFB task(s)	EF task(s)	*r*
Abrams et al. (2014)	UK	6.75	62	Sullivan et al. (1994)	Dot Span forward	.33
					Multiple classification	-.08
Arslan et al. (2017)	Turkey	5.60	41	Perner and Wimmer (1985)	Word Span	.15
				Sullivan et al. (1994)	Listening Span	.51
		7.60			Word Span	.13
					Listening Span	.17
Bibby and McDonald (2005)	Australia	35.90	15	Sullivan et al. (1994)	Digit span	.018
Bock et al. (2015)	USA	8.11	60	Perner and Wimmer (1985)	Digit span	.17
			140	Sullivan et al. (1994)	Location memory	-.07
					Stroop Colour-Word	-.31
					Object multiple classification	.17
					Reading multiple classification	.23
					DCCS	-.31
Bozikas et al. (2011)	Greece	37.40	30	Frith and Corcoran (1996)	Stroop Colour-Word	.42
					Digit span backward	.33
					Digit span forward	.31
					DCCS	.17

Study	Country	Age (years)	*N*	SOFB task(s)	EF task(s)	*r*
					Trail Making	-.27
Brunet (2013)	Canada	8.76	33	Perner and Wimmer (1985)	Stroop Colour-Word	-.11
				Sullivan et al. (1994)	Day-Night	-.41
					Digit span	.60
Caillies et al. (2014)	France	9.00	10	Perner and Wimmer (1985)	Digit span	.42
				Sullivan et al. (1994)	Auditory Attention and statue	.31
Caillies et al. (2012)	France	9.30	10	Perner and Wimmer (1985)	Digit span	.46
					Letter-Number sequencing	.34
					Stroop Colour-Word	.39
					Knock-Tap	-.17
Cassetta et al. (2018)	Canada	9.49	168	Astington et al. (2002)	DCCS	.22
					Stroop Colour-Word	.27
					Digit span	.17
Comay (2009)	Canada	5.73	66	Astington et al. (2002)	Digit span	.30
Ebert (2020)	Germany	5.60	220	Sullivan et al. (1994)	Digit Span	.11
Ford et al. (2017)	Australia	8.45	37	Perner and Wimmer (1985)	Digit span	.19
					Stroop Happy/Sad	-.21
Guajardo and Cartwright (2016)	USA	8.08	31	Perner and Wimmer (1985)	Graphophonological-semantic	-.06
				Sullivan et al. (1994)		
Henry et al. (2017)	France	43.70	33	Rowe et al. (2002)	DCCS	.41

Study	Country	Age (years)	*N*	SOFB task(s)	EF task(s)	*r*
					Digit span	-.15
Hsu and Cheung (2013)	Hong Kong	5.58	54	Sullivan et al. (1994)	their own	.30
Hur et al. (2013)	Korea	23.1	58	Perner and Wimmer (1985)	DCCS	-.11
Huyder and Nilsen (2012)	Canada	6.78	113	Sullivan et al. (1994)	Simon says	.17
			114		DCCS	.32
Kazi et al. (2019)	Greece	6.58	113	Perner and Wimmer (1985)	Word span	.20
					Digit span	.14
					Lexical Stroop	.17
					Matching task	.25
		8.58			Word span	.29
					Digit span	.11
					Lexical Stroop	.21
					Matching task	.13
Kim (2020)	USA	7.19	179	Perner and Wimmer (1985)	Word Span	.44
Lavoie and Talwar (2018)	Canada	8.29	106	Perner and Wimmer (1985)	Digit span	.52
				Sullivan et al. (1994)	Stroop Happy-Sad	.47
Li et al. (2014)	China	10.40	42	Perner and Wimmer (1985)	Day-Night	-.67
				Sullivan et al. (1994)	Digit span	.63
					Plus/Minus	-.38
Nilsen and Valcke (2018)	Canada	7.92	43	Sullivan et al. (1994)	Stroop Red dog/Blue dog	.01
					Digit span	.14

Study	Country	Age (years)	*N*	SOFB task(s)	EF task(s)	*r*
Paine et al. (2018)	UK	6.93	229	Perner and Wimmer (1985)	Response Organization Objects (ROO)	.07
					Visual-Spatial Sequencing (VSS)	.09
Perner et al. (2002)	Austria	5.78	22	Sullivan et al. (1994)	Digit span backward	.41
					Digit span forward	.55
					Go-NoGo	.32
					Tower of London	.51
					Knock-tap	.52
					Statue	.49
					Auditory attention	.53
					Response set	.38
					Visual attention	.55
Peskin et al. (2014)	Canada	8.00	96	Astington et al. (2002)	Go/No-Go	-.05
					Digit span	.32
Raijmakers et al. (2014)	Netherlands	5.50	49	Perner and Wimmer (1985)	Digit span	.19
				Sullivan et al. (1994)		
Rostad and Pexman (2014)	Canada	6.00	80	Perner and Wimmer (1985)	DCCS	.32
				Sullivan et al. (1994)		
Saban-Bezalel et al. (2019)	Israel	11.96	20	Perner and Wimmer (1985)	Trail making A	.05
					Trail Making B	.31
Strasser and del Rio (2013)	Chile	5.54	287	Sullivan et al. (1994)	Pencil-tapping	.09
					Digit span	.23
					Word span	.12

Study	Country	Age (years)	*N*	SOFB task(s)	EF task(s)	*r*
Valle et al. (2015)	Italy	18.46	110	Sullivan et al. (1994)	Listening span	-.02
					Clock test	-.03
Vilenskaya and Lebedeva (2017)	Russia	7.10	30	Perner and Wimmer (1985)	Kogan's task of combining attributes	.13
				Sullivan et al. (1994)	Tower of Hanoi	.08
Williams et al. (2016)	Canada	9.20	79	Perner and Wimmer (1985)	Stroop Colour-Word	.28
				Sullivan et al. (1994)	Digit span	.24