

The Role of Metarepresentation in Preschoolers' Theory of Mind Development

by

Kirsten A. Quistberg
B.A., Queen's University, 2015

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Kirsten Quistberg, 2018
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

The Role of Metarepresentation in Preschoolers' Theory of Mind Development

by

Kirsten A. Quistberg
B.A., Queen's University, 2015

Supervisory Committee

Dr. Ulrich Müller, Department of Psychology
Supervisor

Dr. Erica Woodin, Department of Psychology
Departmental Member

Dr. John Sakaluk, Department of Psychology
Departmental Member

Abstract

Supervisory Committee

Dr. Ulrich Müller, Department of Psychology
Supervisor

Dr. Erica Woodin, Department of Psychology
Departmental Member

Dr. John Sakaluk, Department of Psychology
Departmental Member

The role of metarepresentation in theory of mind development has been hotly debated. On one side of the debate, researchers suggest that theory of mind develops through a domain general change in representational processing of both mental and non-mental representations. On the opposing side, researchers suggest that a unique domain-specific mechanism is required for processing mental representations (i.e., theory of mind). The objective of the current work was to clarify the role of metarepresentation in theory of mind development by examining the relations between children's false belief understanding (mental representations) and non-mental representations (propositional and pictorial representations) understanding. A secondary objective was to investigate the role of conflict inhibition in understanding the representational qualities of beliefs, words, and signs. One-hundred and four three-and-four-year-old children were included in the current analyses. Children's theory of mind understanding was assessed using the false belief (change in location), and the false belief (unexpected contents) tasks. Children's metalinguistic awareness (i.e., propositional representational understanding) was assessed using the Synonym and Homonym Judgment Tasks (see, Doherty & Perner, 1998). False sign tasks were used to assess children's understanding of pictorial representations. Conflict inhibition—the ability to suppress a dominant response in favour of an alternative response—was also measured. Frequentist analyses results showed no significant relationships between false belief understanding and metalinguistic awareness, or false sign understanding.

Bayesian regression analyses revealed that Synonym/Homonym Judgement tasks, False sign tasks, and conflict inhibition supported the null hypothesis of no effect in predicting false belief task performance. These results provide preliminary evidence for unique, domain-specific mechanisms involved in theory of mind development. Results also show that conflict inhibition may be particularly important for success on both the metalinguistic and false sign tasks. Future research should consider looking at the prospective relations between false belief understanding, metalinguistic awareness, and false signs over the preschool period.

Table of Contents

Supervisory Committee	ii
Abstract.....	iii
Table of Contents	v
List of Tables.....	vii
List of Figures	viii
Acknowledgements.....	ix
Dedication	x
Introduction	1
Theory of Mind.....	2
Metarepresentation	6
Metarepresentational mechanisms.....	8
The false sign task.....	15
Metalinguistic Awareness	18
Assessing metalinguistic awareness.....	19
Metalinguistic awareness and false belief understanding.....	21
Related Cognitive Processes: The Role of Executive Function and Language.....	24
Executive function.....	24
Language	27
The Present Study	29
Theory of mind mechanism theory.....	30
Representational processing theory.....	31
Executive function theory.....	32
Representational Content Theory.....	33
Hypotheses	34
Methods	38
Participants.....	38
Measures.....	38
False belief tasks.....	38
Metalinguistic tasks.....	39
Control measures.....	41
Procedure.....	45
Child Development Lab.....	46
Preschools.....	46
Data Analysis Plan.....	46
Data Strategy.....	49
Interpreting bayes factors.....	50
Informed prior distributions.....	51

Results	53
Data Preparation	53
Descriptive Statistics	53
Correlations	57
Zero-order correlations.....	57
Bayesian correlations.	57
Partial correlations.....	61
Partial bayesian correlations.	61
Hierarchical linear regression..	64
Age-related effects.....	68
Performance on alternative naming tasks.	69
Discussion	71
Theory of Mind Mechanism and Domain-Specific Processes.....	71
Evidence against the representational processing theory.....	74
The Role of Conflict inhibition.....	77
Alternative Naming Task Performance	79
Limitations of the Current Study	80
Future Directions	82
Conclusion.....	83
References	85
Appendix A	105
Appendix B	106
Appendix C	107
Appendix D	109
Appendix E	111
Appendix F.....	112
Appendix G	115
Appendix H	116
Appendix I.....	117

List of Tables

<i>Table 1.</i> Proposed hierachical linear regression models	30
<i>Table 2.</i> Mean, range and standard deviations of all variables (N = 104).....	54
<i>Table 3.</i> Mean, range and standard deviations for aggregate scores (N = 104).....	55
<i>Table 4.</i> Independent sample <i>t</i> -test and Baayesian Independent t-test predicting task performance from gender	55
<i>Table 5.</i> Percent of correct responses on controll questions and overall	56
<i>Table 6.</i> Percent of children who identified both items from synonym or homonym pairs in the vocabulary check, and the percent of children who passed the vocabulary check	56
<i>Table 7.</i> Zero-order correlations and Bayesian Correlations between all the false belief, false sign, metalinguistic, executive function and control tasks	59
<i>Table 8.</i> Zero-order correlations and Bayesian Correlations for aggregate scores.....	60
<i>Table 9.</i> Partial correlations and partial Bayesian correlations between all false belief tasks, false sign tasks, and metalinguistic tasks, partialling out conflict inhibition and expressive vocabulary, rrspectively	63
<i>Table 10.</i> Partial correlations and partial Bayesian Correlations between all false belief tasks, false sign tasks, metalinguistic tasks, partialling out conflict inhibition and express vocabulary	64
<i>Table 11.</i> Simple linear regressions predicting task performance from age (in months).....	68
<i>Table 12.</i> A basic heuristic for interpreting Bayes Factor (adapted from Jeffreys, 1961)	51
<i>Table 13.</i> Bayesian linear regression model predicting false belief task performance (null-comparison – including age and expressive vocabulary)	65
<i>Table 14.</i> Bayesian linear regression model predicting false belief performance (null-comparison)	66
<i>Table 15.</i> Bayesian linear regression model predicting false belief task performanc (best possiblemodel comparison)	67
<i>Table 16.</i> Bayesian linear regressions predicting task performance from age (in months).....	66

List of Figures

<i>Figure 1.</i> Theory of Mind Mechanism Theory	9
<i>Figure 2.</i> Representational Processing Theory	12
<i>Figure 3.</i> Shared cognitive processes in false belief understanding and metalinguistic awareness	22
<i>Figure 4.</i> Executive Function Theory.....	32
<i>Figure 5.</i> Representational Content Theory.....	34
<i>Figure 6.</i> Predictions regarding the key process involved in false belief understanding by theory.....	36

Acknowledgements

First, I would like to thank you to my supervisor, Dr. Ulrich Mueller, for his expertise and valuable feedback throughout this process. I would also like to thank my committee members, Dr. Erica Woodin for her insightful comments and suggestions, and Dr. John Sakaluk for his assistance with the Bayesian analysis and helpful recommendations. A sincere thank you to the staff, parents, and children at Gingerbread Preschool, Belmont Park Preschool, and Strawberry Vale preschool for their participation and enthusiasm for the project. I am also extremely grateful to the members of the Child Development Lab whose time and effort to schedule and test children are greatly appreciated. Finally, a heartfelt thanks to my family and friends for their ongoing support and encouragement.

Dedication

To Michael,
Thank you for your endless love, support and encouragement. I am so grateful.

The Role of Metarepresentation Preschooler's Theory of Mind Development

Introduction

Theory of mind—understanding that one's own mental states (e.g., beliefs, desires, intentions) differ from others' (Wellman, Cross & Watson, 2001)—is critical to social competence (for review see, Imuta, Henry, Slaughter, Selcuk & Ruffman, 2016). However, the processes which facilitate theory of mind development remain unclear. Perner (1991) proposed that a conceptual change in understanding representations—specifically understanding mental states as representative of the world—contributes to the emergence of theory of mind. This metarepresentational understanding is particularly important for theory of mind development because it involves simultaneously representing one's own thoughts, while attending to others' mental states (representations). Yet, the central question that remains unresolved is whether theory of mind relies on domain-specific metarepresentational processes specific to mental state attribution, or domain-general advancements in representational processing. The domain-specific account argues the development of theory of mind relies on 'Theory of Mind Mechanism': a specific cognitive mechanism required for understanding mental states (Leslie, 1987; 1994). On the domain-general side of the debate—the Representational Processing Theory—suggests that reasoning about representations in both mental and non-mental domains relies on the same underlying development in representational processes (Perner, 1991; Suddendorf, 1999). In other words, understanding mental representations may extend to other general non-mental representational capacities, including understanding pictorial representations (e.g., signs, labels) and metalinguistic awareness: involving the understanding of language as a flexible, representational system that conveys meaning through the specific structure of words (e.g., speech sounds), and the association amongst words (e.g., word order). Alternative accounts

suggest that the associations between understanding mental and non-mental representations is not attributed to metarepresentation, but rather a variety of other high-order cognitive processes, including perspective taking skills, executive function and language. The current study aims to clarify the role of metarepresentation and other higher-order cognitive processes in the emergence of theory of mind in preschoolers. First, I will discuss theory of mind development and metarepresentational processes. Next, I will review previous evidence for domain-general and domain-specific metarepresentational mechanisms in false belief understanding. Third, I will discuss the relations between false belief development and non-mental representational tasks (e.g., metalinguistic and false sign tasks), as well as executive function and language. I will conclude with the hypotheses of the current work.

Theory of Mind

The ‘theory of mind’ concept was originally derived from comparative psychology research focused on determining if chimpanzees have a theory of other minds (Premack & Woodruff, 1978). Since this seminal work, extensive research has been dedicated to theory of mind. Today, theory of mind is understood as the appreciation of mental states (e.g., beliefs, desires, intentions) as person-specific attributes that motivate behaviour (Saracho, 2014a). Predicting and explaining human behaviours based on inferences about others’ abstract mental states is the most essential application of theory of mind understanding (Sabbagh, Benson, Kuhlmeier, 2013). This inferential processing allows one to intuit that given that a belief is true, and a desire is present, the action will likely follow. For example, if ‘I *want* a cookie’ and ‘I *believe* the cookie is in the cupboard’, then I will likely retrieve the cookie from the cupboard. The capacity to infer others’ mental states and the ability to use these mental states to make actions intelligible is necessary for successful social interactions.

There are a variety of theories used to explain the general method children use to acquire an understanding of other minds (for a review see Carpendale & Lewis, 2015). Perner (1995, 1999) argued children come to understand the concept of beliefs (and other mental states) *as-representations* of the world. From this theoretical framework it is proposed that during the preschool period children develop a representational theory of mind: an explicit understanding of mental states as person-specific representations of the world (Sabbagh & Callanan, 1998; Swiatczak, 2011). Over the course of development children's understanding of other minds becomes increasingly complex (Carlson et al., 2013).

Theory of mind understanding undergoes major changes in early life (Carlson et al., 2013). Already in infancy there may be an implicit understanding of mental states as assessed by looking time tasks (Baillargeon, Scott & He, 2010; Buttelmann, Carpenter & Tomasello, 2009; Buttelmann, Over, Carpenter & Tomasello, 2014; Onishi & Baillargeon, 2005; Southgate, Chevallier & Csibra, 2010; Southgate & Vennetti, 2014; albeit controversial, see meta-analysis by Heyes, 2014). By two to three years, children are able to identify basic emotions, intentions, and desires in themselves and in others (Wellman, 2002; Wellman, Cross & Watson, 2001). By four years, children typically develop an understanding of false beliefs: the recognition that others may hold false beliefs about the world (Astington, 1993). By five years children have a firm, adult-like understanding of mind, namely that all individuals represent world in their minds, and that these mental representations guide behaviours (even if they do not match reality) (Milligan, Astington, & Dack, 2007). There seems to be an orderly progression in preschoolers' performance on false belief tasks (see scale, Wellman & Liu, 2004), and cross-cultural research suggests that theory of mind is acquired in a similar fashion across cultures (Duh et al., 2016; Kuntoro, Saraswati, Peterson & Slaughter, 2013; Liu, Wellman, Tardif & Sabbagh, 2008).

However, contextual factors may lead to slight variations in the order in which different aspects of theory of mind are acquired. Although theory of mind understanding continues to develop into adolescence (Pillow, 1999), the most significant advancements arguably occur during the preschool years.

False-belief understanding. One hallmark of representational theory of mind development during the preschool period is false belief understanding. False belief understanding requires the knowledge that although mental states often represent reality, they are ultimately distinct from reality, and therefore can be false (Sabbagh, Bowman, Evraire, & Ito, 2009; Wellman, Cross, & Watson, 2001). Two types of false belief task—the location-change and unexpected contents tasks—are often used to evaluate young children’s false belief understanding.

Wimmer and Perner (1983) developed the original false belief location-change task. In the location-change false belief task, children are told a story about a character called Maxi. In the story, Maxi puts his chocolate into the cupboard and then leaves the room. However, in Maxi’s absence the chocolate is moved to the drawer. When Maxi returns the children are asked the false belief question: “where will Maxi look for his chocolate: in the cupboard, or in the drawer?” and a reality control-question: “where is Maxi’s chocolate really?” In order to pass this false belief task the child must recognize that Maxi will act on his (false) belief (i.e., look in the cupboard) as opposed to reality (i.e., look in the drawer).

Another widely used task to measure false belief understanding is the unexpected contents false belief task. In this task, children are shown a familiar candy box (e.g., a Smarties box). Before the box is opened the children are asked what they think is in the box (e.g., Smarties or candies). The box is then opened to reveal unexpected contents (e.g., stickers) and is closed

once again (Gopnik & Astington, 1989). Then a new character, Mickey Mouse, is introduced. Children are told that Mickey has never seen inside the box, and then are asked a false belief question: “what will Mickey Mouse think is the box?” and a representational change questions: “what did you think was in the box before we opened it?” The unexpected contents tasks assess children’s understanding of their own false belief (representational change) and the false beliefs of others (e.g., Mickey Mouse). The false belief location and false belief contents task are typically correlated (Müller, Miller, Michalczyk & Karapinka, 2007) and children typically pass these tasks around four-years of age (Wellman et al., 2001). False-belief tasks require the differentiation between one’s own true belief and a character’s false belief, and the understanding that beliefs (and other mental representations) influence behavior.

Implications of theory of mind development. Progress on theory of mind tasks—demonstrating an increasingly complex understandings of mental life—has important implications for socio-cognitive functioning (Tomasello, 2009). A recent meta-analysis of 76 studies (6,432 children) found a significant, yet small in magnitude ($r = .19$) mean effect size for the association between theory of mind and a variety of prosocial behaviours including helping, cooperating and comforting (Imuta, Henry, Slaughter, Selcuk & Ruffman, 2016). False belief mastery has also been found to predict a variety of other positive social and cognitive outcomes including more positive peer relationships (see meta-analysis, Slaughter, Imuta, Peterson & Henry, 2015), more socially apt interactions with peers (Dunn, Cutting, & Demetriou, 2000), higher teacher-rated social competence (Peterson, Slaughter, & Paynter, 2007), improved social communication, and higher academic achievement (Brennan, Galati, & Kuhlen, 2010). Children who develop strong theory of mind skills, therefore, generally demonstrate improved social competence, are more accepted by peers, and do better in school. Conversely, atypically

developing individuals tend to have difficulty with false belief tasks and show specific impairments in social cognition. For example, individuals with Autism Spectrum Disorder (Baron, Leslie & Firth, 1985) and schizophrenia (see meta-analysis, Sprong, Schothorst, Vos, Hox, & Van Engeland, 2007) show significant difficulty with false belief reasoning. Given the significance of false belief understanding for adaptive social-cognitive functioning it is essential that research determines which factors facilitate theory of mind emergence, how theory of mind understanding changes over development, and what contributes to individual differences in mental state understanding (Carlson, Koenig & Harms, 2013). Despite the extensive research devoted to theory of mind, the specific role and function of various cognitive processes in theory of mind development in preschool children remains ill-defined.

Metarepresentation

Metarepresentation may have an important role in the facilitation of theory of mind development and is considered an important acquisition in representational development. According to Perner (1991) there are three levels of representational understanding: primary representation, secondary representation, and meta-representation. ‘Primary’ representation refers to representing the world as it is; for example, the statement “the sun is shining” is a primary representation because it reflects one’s current experience of reality. ‘Secondary’ representations, in contrast, refer to the representation of many possible, or hypothetical situations. For example, the statement, “I believe that the sun is shining” is a secondary representation because this statement may or may not represent reality. Metarepresentation is “a representation of a representation as a representation” (Perner 1991, p. 23). For example, if Suzy says to Lucy “I believe the sun is shining” for Lucy to comprehend this statement she must represent Suzy’s belief (i.e., that it is currently sunny) as a representation that may, or may not be

true. Developing metarepresentational understanding involves the ability to create these higher-order representation (meta-representations) of a lower-order representation (Wilson, 2000). At the foundational level, metarepresentation requires the ability to differentiate between the “sense” and the “referent”. For Frege, the *sense* is the way that a situation is being represented and the *referent* is the ‘truth’ of the situation itself (Frege, 1948). In other words, metarepresentation requires the distinction between what is represented (referent) and how it is represented (sense). This distinction may be particularly important for theory of mind reasoning: understanding that how others represent the world may differ from one’s own representations (or reality).

The representational theory of mind is inherently metarepresentational: it requires the ability to represent others’ (mental) representations as representations (Leslie, 1987). The false belief tasks tap into children’s ability to evaluate other’s misrepresentations of reality, and judge future behaviours accordingly. To pass the location-change false belief task (e.g., Maxi Task described previously) children must acknowledge that the protagonist’s *sense* of the location of the object does not align with the true location of the *referent* (Doherty, 2000). False belief understanding requires the metarepresentational capacity to represent others’ representations (false beliefs) as a representation. Therefore, it is proposed that three- and four-year-old’s struggle with false belief tasks because they are unable to represent others’ mental representations as unique, individualistic representations of the world (Sabbagh & Callanan, 1998).

Research suggests that during the preschool years children typically develop metarepresentational understanding (Juan & Astington, 2012); yet there is no theoretical consensus on how metarepresentational understanding is acquired and thereby contributes to

theory of mind understanding. The central question that remains unanswered is: does theory of mind development rely on a selective mechanism for processing mental states, or general advancements in metarepresentational understanding?

Metarepresentational mechanisms. Over the last two decades the role of metarepresentation in the development of theory of mind has been widely debated (Leekman, Perner, Healey & Sewell, 2008). On one side, some researchers contend that theory of mind relies on a *domain-specific* process for understanding mental states above and beyond general metarepresentational reasoning (Cohen & German, 2010; Frith & Frith, 2003; Leslie, 1994; Leslie, Friedman, & German, 2004; Leslie & Thaiss, 1992; Saxe et al., 2004; Scholl & Leslie, 1999; Scott & Baillargeon, 2009). The ‘Theory of Mind Mechanism’ is the term often used to describe this position (Leslie, 1987; 1994). By contrast, proponents of *domain-general* theories contend that children’s understanding of mental representations (e.g., beliefs, desires, etc.) relies on a more general conceptual change in understanding representations (not specific to mental states) (Leslie, 1987, 1994; Perner, 1991; Stone & Gerrans, 2006; Suddendorf, 1999; Suddendorf & Whiten, 2001; Zaitchik, 1990). The ‘Representational Processing theory’ is often used to describe the domain-general position (Perner, 1998; Suddendorf, 1999). There is evidence to support both positions.

Domain-specific: Theory of Mind Mechanism.

Proponents of the theory of Theory of Mind Mechanism argue that theory of mind develops via an innate metarepresentational capacity specific to processing mental representations (Scholl & Leslie, 1999; Stone & Gerrans, 2006). Within this framework, theory of mind is understood as product of a unique domain-specific cognitive mechanism exclusively dedicated to mental state reasoning. A domain-specific mechanism refers to a process which is

only applied to a certain type of information (e.g., mental representations) (Fodor, 2000). In this conceptualization, other higher-order cognitive processes (language, and so on) might be involved in the expression of theory of mind because theory of mind tasks make certain processing demands. For example, language skills are required to understand the false belief narrative and respond to the questions appropriately. However, the key assertion made by proponents of the Theory of Mind Mechanism theory is that these secondary cognitive demands of theory of mind tasks are not crucial to the concept of theory of mind. Instead, theory of mind relies on the domain-specific processes which are only applied to understanding mental states (Cohen, Sasaki & German, 2015). The Theory of Mind Mechanism theory therefore implies that children acquire theory of mind via domain-specific process whose only function is to reason about mental state attribution. See Figure 1 for a visual depiction of the Theory of Mind Mechanism Theory.

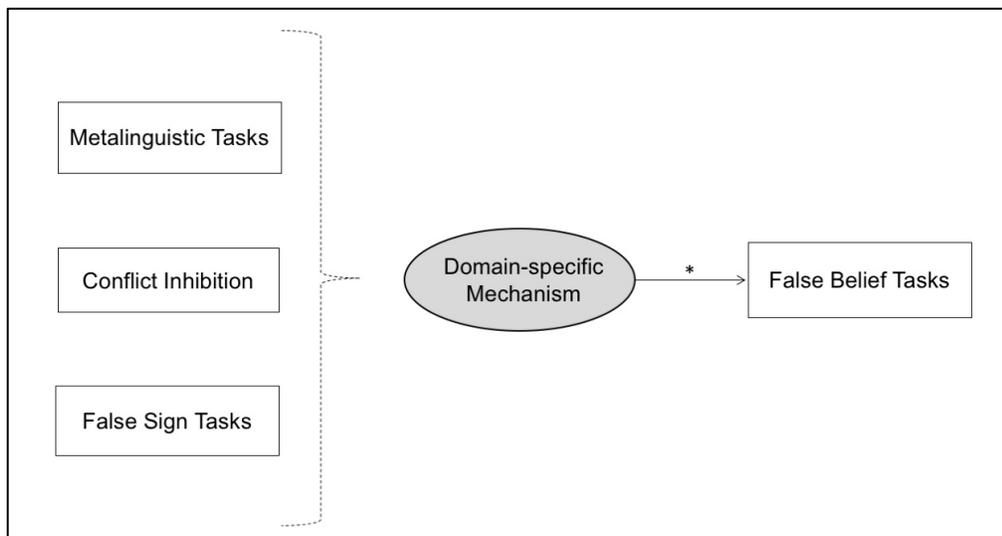


Figure 1. Theory of Mind Mechanism Theory.

Theory of Mind Mechanism: Evidence.

Evidence for the Theory of Mind Mechanism theory comes from performance on the false photograph task, neuroimaging studies, and research with adults. The false photograph task is designed to closely resemble the structure of the location-change false belief task. In the false photograph task, a character (e.g., a doll) “takes” a picture of an object in a certain location with a Polaroid camera; however while the photo is developing the object is moved to an alternative location (Zaitchik, 1990). The photograph then becomes a ‘misrepresentation’ of the location of the object. After hearing this story, children are asked “where is the object in the picture” (test question), and “where is the object really?” (control question). The ability to pass false photograph tasks closely resembles the developmental trajectory of a traditional false belief task (around 4 to 5 years) (Leekam & Perner, 1991; Leslie & Thaiss, 1992; Zaitchik, 1990). However, intriguingly, children with Autism tend to pass the false-photograph task at the same age as typically developing children, yet consistently fail the false belief task (Leekam & Perner, 1991; Leslie & Thais, 1992). Since these tasks have similar metarepresentational demands these results suggest that children with autism have a specific deficit in reasoning about mental representations (as opposed to a general metarepresentational deficit) (see, Leekman & Perner, 1991).

The second line of evidence for domain-specific representational mechanisms for theory of mind is drawn from neuro-imaging studies. Many functional imaging studies (magnetic resonance imaging) have investigated the brain basis of theory of mind. A review of this literature suggests that there are several “core [neural] regions” involved in theory of mind reasoning (Carrington & Bailey, 2009). In adults, the bilateral temporal-parietal junction, precuneus, and medial prefrontal cortex are active when reasoning about false beliefs, rather than

false-photos (Saxe and Kanwisher, 2003). Similarly, ventrolateral prefrontal cortex is activated when completing mental representation tasks (i.e., false belief task), but not when completing non-mental representation tasks (i.e., false photo task) (Hartwright, Apperly & Hansen, 2015). The activation of these regions appears to be specific to reasoning about mental representations. Similarly, in child samples, several studies have identified the unique role of the right-temporal parietal junction (rPTJ) in reasoning about false-beliefs (Sabbagh & Taylor, 2000; Sabbagh, Bowman, Evraire & Ito 2009; Saxe & Wexler, 2005). The identification of these specific neural regions devoted to theory of mind reasoning provides further evidence for a Theory of Mind Mechanism.

Lastly, evidence for domain-specific processes that comprises a ‘theory of mind system’ has been found in an adult population. Specifically, a study by Cohen, Sasaski and German (2014) indicates that adults have a unique advantage for processing mental representations in comparison to other non-mental representations (e.g., linguistic, signs etc.) when task demands are controlled for. However, there is no research, to our knowledge, that has replicated these findings in child or adult samples.

Domain-general: Representational processing theory.

On the domain-general side of the debate, theorists argue that a general cognitive mechanism for understanding representation and metarepresentation is required for theory of mind development. This position is often described as the ‘Representational Processing Theory’ (Perner, 1991; Suddendorf, 1999). Proponents of the Representational Processing Theory argue that theory of mind develops via a general conceptual change in understanding representations, not specific to belief-representation. From this perspective, children have difficulty understanding mental states “not because they are mental but because they are representations”

(Zaitchik, 1990, p. 60). The Representational Processing Theory rests on the notion that both mental and non-mental representations require the ability differentiate between the (1) referent (i.e., the thing itself) and (2) the sense (the way it is represented) (Frege, 1948; Perner, 1991). For example, understanding that a woman can be both a mother and a teacher requires the ability to differentiate between *what* is being represented (i.e., the woman), and *how* she is being represented (as either a mother or a teacher). The Representational Processing Theory suggests that theory of mind develops as a result of general advancements in representational understanding. See Figure 2 for a visual depiction of the Representational Processing Theory.

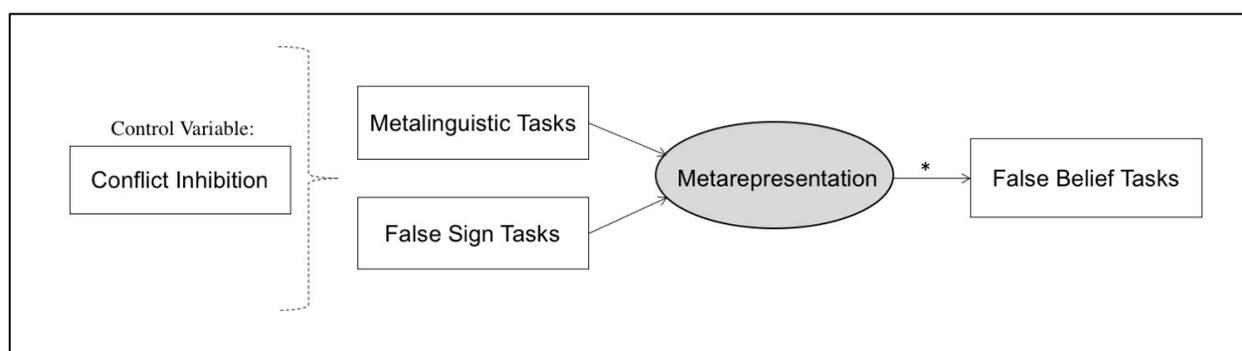


Figure 2. Representational Processing Theory.

Representational processing theory: Evidence.

Support for the Representational Processing Theory is often presented as counter-evidence to the domain-specificity hypotheses previously discussed. A key piece of evidence in favour of the Theory of Mind Mechanism theory is the performance of children with autism on the false photograph task. However, there are important reasons why the false photograph task may be an inadequate measure of (meta)representational understanding. The central problem with the false photograph task is often referred to as the “falseness problem”. The “falseness problem” refers to the photograph in the task, which is not a false representation, but rather represents a “true” state of affairs in the past (Perner & Leekman, 2008; Sabbagh, Moses, &

Shiverick, 2006). For instance, a person does not look at a picture of themselves on vacation a year ago and presume that this is a false representation. In other words, photographs represent the world correctly as it was, not as a current misrepresentation. Additionally, in contrast to Zaitchik's (1990) argument, false photograph and false belief tasks are not structurally equivalent. False belief tasks involve more abstract representations and higher executive function demands than the false photograph task (Müller, Zelazo, Imrisek, 2005; Sabbagh, Moses, Shiverick, 2006). Photographs physically instantiate the representation (the photo itself) that does not need to be held in mind nor inhibited. False-belief tasks also make the additional demand of understanding propositional statements (i.e., assessing the 'truthfulness' of a statement), which the false photograph task does not. As a result, the false photograph task does not rely on the ability to differentiate between what is being represented and how it is being represented, and therefore is not an appropriate measure of metarepresentational understanding (Leekam Perner, Healey & Sewell, 2008). Using evidence from the performance of children with autism on a false photograph task as evidence for a domain-specific mechanism for understanding mental representations is problematic.

The evidence from neuroimaging studies defining the specific neural regions involved in false belief reasoning has also been contested. Lin and colleagues (2018) analyzed the neural correlates for processes involved in theory of mind—namely, social concept representation and retrieval, domain-general semantic integration, and domain-specific integration of social semantic contents—and found diffuse activation of these regions for all processes. Similarly a variety of studies have demonstrated there is a lack of specificity in the neural regions activated while completing theory of mind tasks in comparisons to other related tasks (e.g., false-photo, false-sign, etc.) (Perner & Leekam, 2008; Stone & Gerrans, 2006). Lesions in the left TPJ have

also been found to cause difficulty in reasoning about false beliefs as well as false photographs (Apperly, Samson, Chiavarino, Bickerton, & Humphreys, 2007). Furthermore, at a more fundamental level, simply because a neural circuit is involved in false belief reasoning does not imply that this circuit is specific to that particular task (Stone & Gerrans, 2006). Collectively, these findings suggest that these neural regions are required mental state reasoning are diverse, and therefore evidence for a neural region specialized for theory of mind reasoning is not supported. Indeed, the neural regions recruited for reasoning about representations do not appear to be specific to mental representations providing further evidence for the Representational Processing Theory.

Another line of evidence for the Representational Processing theory is children's performance on mental and non-mental representational tasks. Empirical support for preschoolers' difficulty on both false belief tasks and other non-mental tasks is consistently identified in the literature (Doherty & Perner, 1998; Doherty, 2000; Leekman, Perner, Healey & Sewell, 2008; Perner, 1991; Sabbagh & Callanan, 1998; Swiatczak, 2011). These findings provide preliminary evidence for a general conceptual change in representational processing. However, to our knowledge, no study to-date has compared children's performance on the false belief task to different types of non-mental representations, namely symbolic representations (false signs) and linguistic representations (e.g., metalinguistic awareness). The current study will compare children's performance on false belief tasks to both non-mental general representations (false signs) and linguistic representations (metalinguistic awareness).

Non-mental (or physical) representational tasks are design to test children's understanding of the representational nature of pictures, signs, maps, models, and words. Analyzing and comparing children's performance on tasks that assess non-mental representations

and mental representations (i.e., false beliefs) provides an opportunity to determine their metarepresentational basis (domain-specific or domain general). Two tasks often used to evaluate non-mental representations include the false-sign tasks (pictorial representations), and alternative naming tasks (linguistic representations). Although the age at which children pass non-mental and mental representation tasks is generally similar (around four years of age) (Iao, Leekam, Perner & McConachie, 2011; Leekam, Perner, Healey, & Sewell, 2008; Sabbagh, Moses & Shiverick, 2006; Wellman et al., 2001), *why* children perform similarly on these tasks is less clear.

The false sign task.

The False Sign task, originally introduced by Parkin and Perner (2004), is designed to assess whether children's ability to pass false belief tasks of more general representational abilities. The False Sign task, like the false belief task, has two variants: the location and contents version. In the False-Sign location version children are shown an arrow which is used to signify the location of an object/person. Then, the object/person is moved to a new location, but the arrow remains pointing to the original location. The children are then asked a control question: "where is the object/person really?" and a representational change question: "Where does the arrow say the object/person is?" (Sabbagh, Moses & Shiverick, 2006; Iao, Leekam, Perner & McConnachie, 2011). In the False sign contents version, children are shown a box with a label (e.g., a Band-aid box), and then shown the true contents of the box (e.g., crayons). Children are then asked a control question: "what is in the box really?" and a representational change question: "what does the label show is in the box?" (Sabbagh, et al., 2006; Iao et al., 2011). False sign tasks are elegantly designed to mirror the structure of traditional false belief tasks and evaluate (non-mental) representational understanding.

Significance of false signs.

False-sign tasks provide a unique and important resource for assessing the role of representations in a non-mental domain. False-sign tasks were developed as a means of solving the “falseness problem” of the false photograph task (i.e., a photo could represent a ‘true’ state of affairs in the past). Unlike the false photograph task, the false sign task incorporates a genuine false representation (e.g., arrow pointing in the wrong direction) of the current state in the same way as a false belief. Additionally, in line with the false belief task, the false sign task involves similar cognitive demands, namely executive function—higher-order cognitive processes involved in problem-solving and goal-directed behaviour—and metarepresentation. Executive function is required to inhibit a ‘true’ representation in favour of the non-dominant ‘false’ representation. Executive function significantly predicts performance on both the false sign and the false belief tasks (Sabbagh, Moses, & Shiverick, 2006). Metarepresentational understanding is also required to pass False Sign tasks because, like False belief tasks, it requires the understanding of and differentiation between what is represented (i.e., true location of the object), and how a representation is understood (i.e., perspective on representation; the arrow) (Parker & Perner, 2004). The current research utilizes the false sign task to explore children’s understanding of non-mental representations, using a task version which makes similar task demands as the traditional false belief tasks.

False signs & false beliefs.

Significant associations between false-sign and false belief task performance have been consistently identified in the literature (Iao, Leekman, Perner & McConachie, 2011; Leekam, Perner, Healey, & Sewell, 2008; Sabbagh, Moses, and Shiverick, 2006). Iao and colleagues (2011) suggest that the robust correlation between False sign and False belief task performance is

due to a conceptual change in representational understanding, not limited to mental states. Additional support for the domain general hypotheses comes from research that shows that children with autism have difficulties on both false belief and false sign tasks (Bowler, Briskman, Gurvidi & Fornells-Ambrojo, 2005). Therefore, children with autism may have a ‘metarepresentational deficit’ as opposed to a domain-specific impairment in mental state reasoning (as suggested by research using the false photograph task). Taken together, these findings suggest that a domain-general component may therefore be necessary to update misrepresentations.

Importantly, the content of the representations in the false sign and false belief tasks differ. The representational content of false sign tasks is pictorial—there is a physical instantiation (e.g., arrow, label) of the misrepresentation which the child can readily access (Sabbagh, Moses, Shiverick, 2006). However, the content of the representations in false belief tasks are propositional. The propositional content of the representations refers to the ‘sentence-like’ format of the False Belief tasks. In other words, understanding mental representations via this task is contingent on the ability to reason about these states via a complex linguistic format. Cohen, Sasaski and German (2016) refer to the difference in the content of the physical content of the representations in false sign tasks and the propositional representational content of the false belief task as the “content problem”. Therefore, in addition to assessing children’s general understanding of pictorial representations it is important to evaluate how children perform non-mental tasks with representational content similar to false beliefs. Introducing other non-mental representational tasks, namely metalinguistic tasks, which have propositional content to their representational demands may resolve this issue.

Metalinguistic Awareness

Metalinguistic awareness is defined as "the ability to reflect upon and manipulate the structural features of spoken language, treating language itself as an object of thought, as opposed to simply using the language system to comprehend and produce sentences"(Tunmer & Herriman, 1984, p.12). Metalinguistic awareness is the understanding of language as a flexible, representational system, and requires the ability to "represent the relationship between linguistic form and what it represents" (Doherty, 2000, p. 387). Metalinguistic abilities are typically described in using the various aspects of language, including "phonemic awareness (i.e. understanding of the sounds associated with speech), syntax awareness (i.e. grammatical judgments), word/semantic awareness (i.e. how words are formed and what they represent), and pragmatic awareness (i.e. metalinguistic communication in social conversation)" (Solesa-Grijak, 2011, p. 44). Metalinguistic awareness emerges during early childhood, and children's understanding of metalinguistic concepts becomes increasingly complex over the course of development (Chaney, 1992; Edwards & Kirkpatrick, 1999; Smith & Tager-Flusberg, 1982). The advancement of metalinguistic abilities have important implications for children's early reading and spelling skills (Carlson, Jenkins, Li & Brownell, 2013; Roman, Kirby, Parrila, Wade-Woolley, & Deacon, 2009). Therefore, it is also important to determine the process which facilitate its emergence, specifically the role of metarepresentational development.

Metalinguistic awareness relies on metarepresentational processes because it involves representing a [linguistic] representation as representations. As children develop a metalinguistic understanding they begin to develop the ability to separate a word from the object it represents (Pan & Gleason, 1997) and appreciate that words can be symbols for things that do and do not exist in the world (i.e., the word 'is' does not refer to an object in the world) (Papandropoulou &

Sinclair, 1974). For the development of semantic awareness—understanding the representational qualities of words—metarepresentational processes may be especially important. Semantic awareness is often evaluated using alternative naming tasks, specifically the Synonym Judgement task and the Homonym judgement task. These tasks require the ability to differentiate between what is being represented (e.g., picture or word) and how it is represented.

Assessing metalinguistic awareness.

Alternative naming tasks are often used to test different elements of children's semantic awareness. In order to pass alternative naming tasks, children must understand that a word is an abstract representational vehicle that is not inherently tied to a single object or phenomena. For example, understanding that the words 'street' and 'road' refer to the same entity (synonym understanding), and conversely, the word 'bat' can refer both to a flying animal and to a piece of sports equipment (homonym understanding). The representations in these tasks have a propositional content: children must reason about the truthfulness of a statement by thinking about words as symbols of ideas. Consistent with the developmental trajectory of metalinguistic development more broadly, children tend to master the synonym and homonym task around four years of age (Doherty & Perner, 1998; Doherty, 2000; Garnham et al., 2002; Perner & Leahy, 2015).

Synonyms—two words that refer to the same thing (e.g., bunny/rabbit)—require the metarepresentational capacity to understand an object (i.e., what is being represented) can be represented by two different word forms (i.e., how it is represented). The synonym judgment task, originally developed by Doherty and Perner (1998), was designed for preschool-aged children. In the task, children are shown a page with four images. To begin, a vocabulary check is conducted to ensure the child knows the meaning of terms used. Next, the child is introduced

to a puppet, and is told that it is the puppet's job to say the "other name" for the object. The child is then shown the pictures again and asked to name the object (e.g., "bunny"). The puppet responds with either the correct synonym pair (e.g., rabbit) or the incorrect response (e.g., spoon), and then the child is asked: "is that what the puppet was supposed to say?" The synonym judgement task requires the ability to assess whether a linguistic representational propositional statement is judged to be correct or incorrect.

Homonyms—understanding that one word can represent two different entities—similarly require metarepresentational understanding. The Homonym Judgement task (Doherty, 2000) is designed to parallel the structure of the Synonym Judgement Task. To begin a vocabulary check is conducted to ensure the child knows the meaning of the terms being used. Next, a puppet is introduced, and the child is told it is the puppet's job to point to the other picture with the same name. The child is instructed to point to a named object, such as 'bat' (e.g., baseball bat). Next, a puppet either correctly point to the other object with the same name (e.g., flying animal bat), or incorrectly points to an alternative item (e.g., bus). The child is then asked, "Is that what the puppet should've done?" The homonym task requires the understanding that two pictures can be represented by a single linguistic form.

The Colour-colour task is often used as an alternative naming control task (Perner & Leahy, 2015). The colour-colour task is designed to be structurally equivalent to the Synonym Judgement Task and the Homonym Judgement task. In this task, the child is tested on their colour vocabulary. Then, the child is shown a variety of pictures that have two colours (e.g., red and blue socks). Similarly, to the other alternative naming tasks previously described, this task requires children to name one colour and then evaluate if a puppet correctly names the other colour. Importantly, the Colour-Colour task presumably makes conflict inhibition demands, but

does not require metarepresentational skills because it does not require representing the same object in multiple ways (Perner, Stummer, Sprung & Doherty, 2002).

Metalinguistic awareness and false belief understanding.

Over the last few decades, a great deal of research has been dedicated to understanding the relations between metalinguistic awareness and theory of mind (Doherty, 2000; Doherty & Perner, 1998; Perner & Leahy, 2015). This line of research provides evidence for significant correlations between metalinguistic awareness and false belief understanding (Doherty, 2000; Doherty & Perner, 1998; Farrar & Ashwell, 2012; Farrar, Ashwell & Maag, 2005; Hacin, 2016). Children are typically able to pass both the metalinguistic tasks and false belief tasks around four years of age (Edwards & Kirkpatrick, 1999; Wellman, Cross & Watson, 2001). However, the explanation for the strong association between these constructs is still debated. Similar cognitive processes, namely metarepresentational understanding, executive function, and language (see Figure 3). may be important for the development of metalinguistic awareness and false belief understanding; yet, how these cognitive processes contribute to children's success (or failure) on both types of tasks is not well understood. A variety of theories have been put forth to explain this association, including the Representational Processing theory.

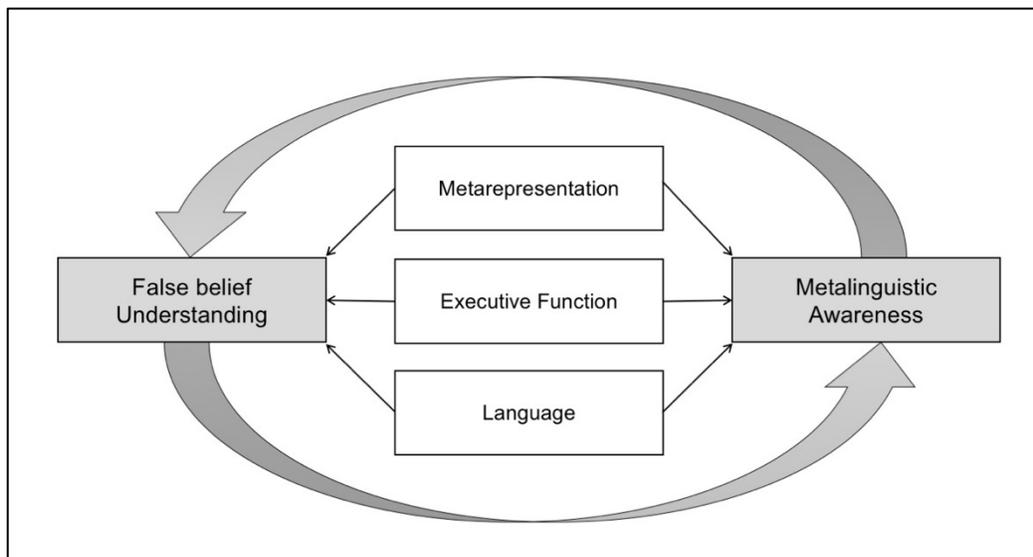


Figure 3. Shared cognitive processes in false belief understanding and metalinguistic awareness.

Proponents of the Representational Processing theory suggests the shared metarepresentational demands of the metalinguistic tasks and false belief tasks explain this relation. As was previously discussed, according to the Representational Processing theory framework, the association between these tasks is due to their shared reliance on the ability to differentiate between a sense (how it is represented), and the referent (the truth of the representation). For instance, the Synonym judgement requires the metarepresentational capacity differentiate between *what* is being represented (e.g., a rabbit; the referent), and *how* it is being represented (e.g., as both a bunny and a rabbit; the sense). Similarly, the false belief task requires the metarepresentational capacity to differentiate between *what* is being represented (i.e., contents of the box; the referent) and *how* the contents of the box are being represented (the true versus assumed contents; the sense). Doherty and Perner (1998) contend that by four years of age children begin to understand the representational nature of language and mental states and this conceptual change in representational understanding explains the association between false belief and metalinguistic task performance.

Additionally, the Representational Processing framework emphasizes the role of perspective taking in metarepresentation understanding. In recent years, Perner and colleagues (Perner, & Leahy, 2016; Perner, Huemer & Leahy, 2015) have utilized a ‘mental files’ computer based analogy to clarify the cognitive processes involved in representing multiple perspectives (or senses) on a representation. The mental files conceptualization suggests there is a specific cognitive mechanisms for storing and attaching representational qualities to a referent (Losada, 2016; Perner & Leahy, 2015). For example, conceptualizing a person involves storing and anchoring a variety of representational properties (or mental files) to that person (or referent), such as ‘friend’, ‘tall’, ‘kind’, ‘blonde’ and so on. Vicarious mental files—representing (or creating mental files) for other representations (mental files)—may be particularly important for metarepresentation. With respect to metalinguistic awareness, proponents of the Mental Files Theory would suggest that success on metalinguistic tasks requires the ability to represent words from a variety of perspectives. For example, the Synonym Judgment Task requires children to not only represent the object from their own perspective (e.g., as a bunny), but also represent another agent’s perspective (i.e., as a rabbit). In other words, children must store a mental file for their representation (e.g., bunny file), as well as a vicarious mental file for the other agent’s perspective (e.g., rabbit file). Similarly, in false belief reasoning, children must hold a mental file for their own perspective (e.g., a file for the true location of an object), as well as a vicarious mental file for the other agent’s perspective (e.g., a file for the false representation of the location of the object). The ability to store and anchor multiple files to a single object or idea not only relies on representational capacities, it requires the ability to represent others’ perspective on a representation (metarepresentation). In other words, the ability to “bracket” one’s own perspective on a representation (mental file) and shift to an alternative perspective on the referent

(vicarious file) is essential to metarepresentation (Farrah & Ashwell, 2012). Therefore, champions of the Representational Processing Theory suggest that shared metarepresentational demands explains the association between the metalinguistic and false belief task performance.

Related Cognitive Processes: The Role of Executive Function and Language

Given the breadth of the research previously described, it is clear that there is much more to the false belief tasks than theory of mind alone (Bloom & German, 2000). The focus of the current work is to determine the role of metarepresentation in false belief understanding; however, in order to effectively clarify the role of metarepresentation, other related cognitive processes must be controlled for. Specifically, children's executive function and language skills must be controlled for because of the significant correlation between these skills and metarepresentational tasks. The relations between these processes and performance on the false belief, false sign and alternative naming tasks in preschool children are described below.

Executive function.

Executive function refers to the higher cognitive processes such as working memory, inhibition, and flexibility involved in the control of action, thought, and emotion that underlie goal-directed behaviour (Zelazo & Müller, 2010). Due to the advancements in the development of the prefrontal cortex during the preschool period children's executive function improves significantly over this period (Müller & Kerns, 2015). A great deal of research has been dedicated to analyzing the role of executive function in the development of other higher-order cognitive mechanisms, including false belief understanding, false signs, and metalinguistic awareness.

False belief understanding & executive function.

The role of executive function in false belief understanding development has been a prominent focus of research in developmental psychology. A recent meta-analysis that included 102 studies (representing over 9,994 participants between the ages 3 to 6 years from 15 countries) found a medium to large effect size ($r = .38$) for the relation between false belief understanding and executive function (Devine & Hughes, 2014). An analysis of 10 longitudinal studies supported an asymmetric relation between executive function and false belief understanding suggesting executive function predicts false belief understanding, but not the reverse (Devine & Hughes, 2014). Research focused on three-and-four-year-old children similarly found a significant correlation between executive function and false belief understanding controlling for chronological age, verbal ability and gender (Müller et al., 2012; Müller, Zelazo & Imrisek, 2005). Cross-cultural research has also identified a correlation between false belief understanding and executive function (Oh & Lewis, 2008; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). These results support significant associations in the emergence of executive function and false belief understanding in young children.

False belief understanding & conflict inhibition.

Conflict inhibition, a specific aspect of executive function, may be particularly important for theory of mind reasoning (Carlson, Claxton & Moses, 2015; Carlson & Moses, 2001). Conflict inhibition refers to the ability to inhibit a dominant response and shift to an alternative response while pursuing a goal. Carlson and Moses's (2001) 'Grass/Snow' task is a prototypical measure of conflict inhibition in preschool aged children. The Grass/Snow task involves pointing to a green card when the experimenter says 'snow' and to a white card when the experimenter says 'grass'. Young children find this task particularly difficult because they must inhibit a

dominant response (e.g., white card for snow), and switch to a non-dominant response (e.g., green card). Carlson and Moses (2001) found that conflict inhibition was significantly related to theory of mind development, and therefore suggest inhibition may facilitate false belief understanding.

There are important reasons to infer a relationship between theory of mind and conflict inhibition. First, conflict inhibition and theory of mind may rely on similar neural regions in the frontal lobe (Fletcher et al., 1995); however, it should be noted that more recent evidence has suggested that the dorsal lateral prefrontal cortex is correlated with theory of mind, even after controlling for conflict inhibition (Sabbagh, Bowman, Evaire, & Ito, 2009). Second, false belief task success requires conflict inhibition (inhibiting one's dominant or egocentric perspective), and switching to the other's perspective (Farrah & Ashwell, 2012). Third, and finally, children with Autism Spectrum disorder may show difficulty with both inhibitory control (Christ, Holt, White & Green, 2007) and false belief tasks (Baron-Cohen, Leslie & Frith, 1985). However, it is important to note some research indicates that there are no inhibitory control deficits in children with Autism Spectrum Disorder (see review, Hill, 2004). Thus, conflict inhibition skills may influence false belief task success, and thus should be controlled for when evaluating theory of mind understanding.

Conflict inhibition & false sign understanding.

Conflict inhibition has also been found to predict preschoolers' performance on the false sign task (Iao & Leekam, 2014; Sabbagh, Moses & Shiverick, 2006). Similarly, to the false belief task, in order to pass the false sign task children must inhibit their dominant response (i.e., true location of the object, or the true contents of a box) in favour of the false representation (i.e., arrow pointing in the wrong direction; deceptive label of box). However, no research to-date has

looked at the relation between conflict inhibition and false sign reasoning in the context of other cognitive processes (e.g., metarepresentation, perspective-taking, and language).

Conflict Inhibition & metalinguistic awareness.

Metalinguistic awareness involves the executive function capacity to inhibit a dominant response (e.g., prototypical understanding of words), and shift to an alternative aspect of language (e.g., sound, grammar, or semantics) (Farrar & Ashwell, 2012; van Kleeck, 1982). For example, Doherty's (2000) homonym task (a metalinguistic awareness task) requires the inhibition of one meaning of a word (e.g., bat as flying animal) at the expense of an alternative meaning (e.g., bat as sports equipment). Additionally, conflict inhibition has also been shown to be significantly correlated with children's understanding of rhyming (Farrar & Ashwell, 2012). Rhyming requires the inhibition of the literal meaning of words, which is dominant in everyday conversation, and the shifting to a focus on phonetics (speech sounds). Thus, reflecting on the different properties of words requires the inhibition of one perspective in favour of an alternative perspective on the same linguistic form (or meaning).

Language

Language & false belief understanding.

Research suggests that language plays a significant role in the promotion of false belief understanding. For example, Milligan, Astington, and Dack's (2007) meta-analysis found that across 104 studies ($n = 9,000$) language ability predicts false-belief understanding (but not the reverse) with a moderate effects size ($r = 0.31$; 95% CI) after controlling for chronological age. In this meta-analysis, language ability was operationally defined as general language, semantics, receptive vocabulary, syntax, and memory for complements (Milligan, Astington, & Dack, 2007, p. 634). Astington and Jenkins (1999) propose a unidirectional association between language and

false-belief understanding, such that earlier language ability may lead to improved false belief task performance later in development. Substantial evidence has supported this claim. For example, Hale and Flusberg (2003) demonstrated that language-based interventions improved children's performance on false belief tasks. Additionally, it is important to consider the language demands of most false-belief task when interpreting these results. That is, some theorists see language as an important tool for structuring one's understanding of belief states (see, de Villiers & de Villiers, 2000).

Language & false sign understanding.

Little research has looked at the exact role of language in solving false sign problems. However, Iao and Leekman (2014) found that language (vocabulary knowledge) predicted performance of the false sign task. Given the verbal demands of the task—comprehension of the narrative and questions, eliciting a verbal response—the role of language in false sign tasks is not surprising. The role of language in false sign (representations) will be investigated further in the current work.

Language & metalinguistic understanding.

Metalinguistic awareness develops as language develops. As children learn language they become increasingly aware of the different aspects of language (e.g., grammar, phonetics, semantics etc.) (Flood & Salrus, 1982; Sharpe & Zelazo, 2002). For example, Smith and Tager-Flusberg (1982), found that children's receptive vocabulary (measured by the Peabody Picture Vocabulary Test) and sentence comprehension independently predicted their performance on a variety of metalinguistic tasks after controlling for chronological age. Metalinguistic understanding is part of children's broader language development, and therefore the significant associations between metalinguistic awareness and language are expected.

The Present Study

The central objective of the present study is to evaluate the theoretical frameworks—Theory of Mind Mechanism, Representational Processing Theory, the Executive Function theory, and the Representational Contents theory—proposed to explain the cognitive processes which contribute to the emergence of theory of mind. There is significant evidence for the importance of metarepresentation and executive function in false-belief understanding. However, *how* these cognitive mechanisms contribute to theory of mind development is less clear. By predicting false belief task performance from false sign task performance (non-mental *pictorial* representations), metalinguistic task performance (non-mental *propositional* representations) and conflict inhibition tasks, the role of metarepresentation and executive function in theory of mind understanding will be further clarified. For proposed statistical Models 1, 2, 3 and 4, see Table 1.

Table 1.

Proposed hierarchical linear regression models.

Model 1	Model 2	Model 3	Model 4
Block 1 Age (Months) Expressive vocabulary	Block 1 Age (Months) Expressive vocabulary	Block 1 Age (Months) Expressive vocabulary	Block 1 Age (Months) Expressive vocabulary
Block 2 Age (Months) Expressive vocabulary Inhibitory control tasks	Block 2 Age (Months) Expressive vocabulary Inhibitory control tasks	Block 2 Age (Months) Expressive vocabulary Inhibitory control tasks	Block 2 Age (Months) Expressive vocabulary Inhibitory control tasks
Block 3 Age (Months) Expressive vocabulary Inhibitory control tasks False Signs	Block 3 Age (Months) Expressive vocabulary Inhibitory control tasks False Signs	Block 3 Age (Months) Expressive vocabulary Inhibitory control tasks Metalinguistic Tasks	
Block 4 Age (Months) Expressive vocabulary Inhibitory control tasks False Signs Metalinguistic Tasks			

Theory of mind mechanism theory.

Proponents of the Theory of Mind Mechanism Theory argues that theory of mind develops via an innate cognitive processes metarepresentational process which is unique to mental state understanding (Leslie, 1989). From this theoretical framework, the ability to represent others mental representations (metarepresentation) is a skill specific to theory of mind. All other cognitive processes such as language, executive function, and so on, play an auxiliary role representational reasoning (Cohen, Sasaki & German, 2015). The current study will evaluate this theory by determining whether variance in false-belief task performance would remain after partialling out the effects of metalinguistic task performance (Synonym and Homonym Task),

false sign task performance, conflict inhibition tasks, general language abilities and chronological age (Model 1). If there is no reduction in the amount of variance in false belief task performance then it would be inferred that these processes are not involved in mental state reasoning. The current study does not include all potential cognitive processes that could contribute to false belief reasoning (e.g., parent theory of mind, intelligence quotient etc.). As a result, we are unable to conduct a rigorous test of the domain-specific processes in theory of mind reasoning. However, if false-belief understanding emerges as a unique skill this will be considered preliminary evidence for a potential domain-specific metarepresentational mechanism for theory of mind reasoning.

Representational processing theory.

In contrast to the Theory of Mind Mechanism theory, proponents of the Representational Processing Theory suggest that the ability to reason about mental states relies on a domain-general conceptual change in representational understanding. In other words, all representational processes rely on the same ability to differentiate between what is being represented and how it is being represented (Perner, 1991). According to this framework, metalinguistic task performance will significantly predict false belief task performance, while controlling for age, language, and conflict inhibition tasks (Model 2). Similarly, false sign task performance will significantly predict false belief task performance, while controlling for age, language, and conflict inhibition tasks (Model 3). Further, there will be no significant difference between the fit of Model 2 and Model 3 because these tasks similarly rely on an understanding of metarepresentation. Additionally, this theory predicts the false belief tasks, false sign tasks, and metalinguistic tasks will be highly correlated. Evidence for the Representational Processing Theory would suggest that during the preschool period children experience a conceptual change

in representational understanding, thereby allowing them to solve false belief tasks, false sign tasks, and metalinguistic problems.

Executive function theory.

Proponents of the Executive Function Theory maintains that the relations between these constructs is due to a shared reliance on executive function skills, specifically conflict inhibition (Garnham & Garnham, 2000). False belief understanding (Devine & Hughes, 2014), the synonym and homonym judgement task (Farrar & Ashwell, 2012), and the false sign tasks (Iao & Leekman, 2014; Sabbagh, Moses & Shiverick, 2006) have been found to be significantly correlated with executive function. However, research to-date has not explored the contribution of executive function to all three metarepresentational tasks in one study, nor controlled for executive function in this relationship. See Figure 4 for a visual depiction of the Executive Function theory.

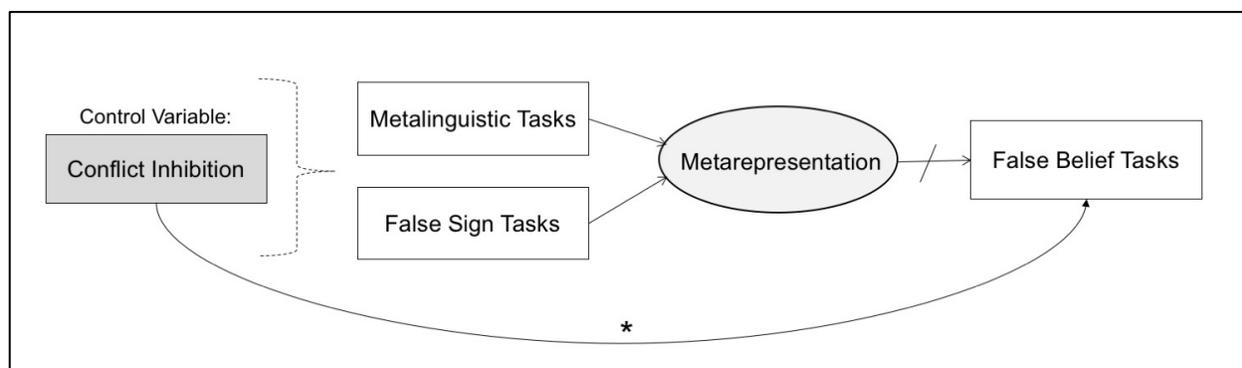


Figure 4. Executive Function Theory.

The hypotheses drawn from the Executive function theory are three-fold. First, the Executive Function Theory framework predicts that conflict inhibition tasks will significantly predict false belief understanding, while controlling for language and chronological age (Model 4). The second prediction that follows from this theory is that both Model 2 and Model 3 will not be significant (because conflict inhibition tasks are controlled for). Finally, performance on

the colour-colour task will not be statistically different than performance on the Synonym or Homonym Judgement tasks, respectively. The colour-colour task is presumed to have the same executive function (conflict inhibition) demands as the Synonym or Homonym Judgement tasks, but not the same representational demands. Therefore, if there is no difference in performance on all three of these alternative naming tasks it would be inferred that the executive function demands of the task, rather than the representational demands, would explain this association.

Representational Content Theory.

As was previously discussed, Cohen, Sasaki and German (2015) emphasize the “content problem” in traditional research on the development of mental and non-mental representations. The content problem refers to the difference in the content of the representations in prototypical tasks used to assess non-mental representational (e.g., signs, photographs, maps), and mental representation tasks (e.g., false belief tasks). Specifically, the former (non-mental representations) involves pictorial representations, or representations with a physical instantiation (e.g., a false sign). By contrast, the latter (mental representations) have a propositional (sentence-like) content that is either true or false. For example, an agent’s knowledge of the location of an object cannot be partially correct—the item (or referent) either is or is not in a given location. Thus, in the current work, introducing the metalinguistic tasks provides an elegant way to assess non-mental representations that have the same (propositional) content as the mental representation tasks, allowing for a more direct assessment of the main research question: is there something unique about mental representations? Thus, I propose a novel prediction which contends that the content of the representation may be particularly important to understanding the development of mental representations. The central tenet of this theory is that metalinguistic tasks will better predict false belief performance, than false sign

tasks because of their shared representational content (see Figure 5). Therefore, from the Representational Content Theory framework, it is hypothesized that metalinguistic tasks will significantly predict false belief task performance, over and above conflict inhibition tasks, language and age (Model 2). Additionally, the Representational Content Theory framework would predict that Model 2 will have a better fit compared to Model 3 (see above). Overall, the Representational Content Theory proposes that the ability to represent linguistic-propositional representations—exemplified by the unique relationship between false belief understanding and metalinguistic awareness—is the key cognitive mechanism involved in theory of mind development.

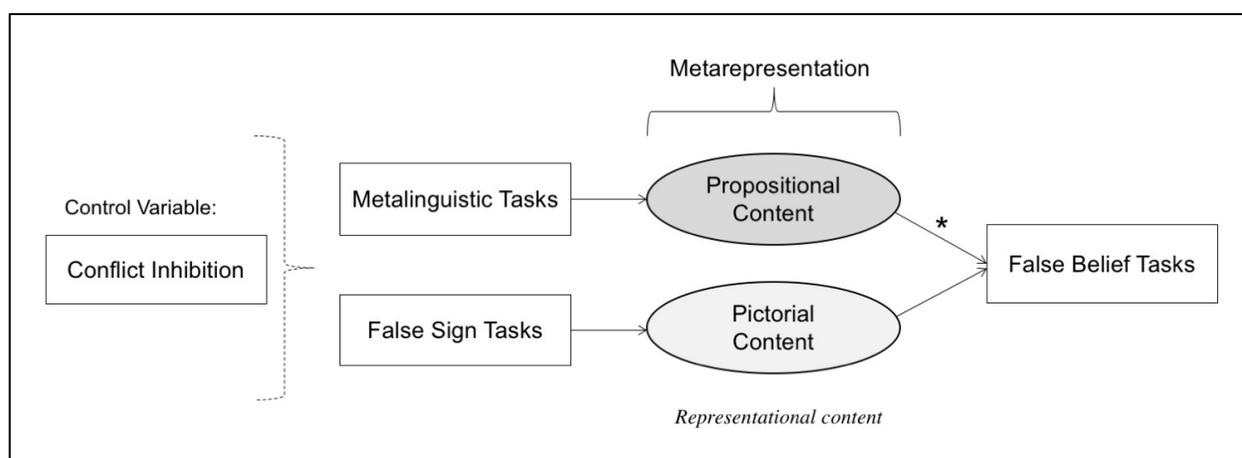


Figure 5. Representational Content Theory.

Hypotheses

I hypothesize that the Theory of Mind Mechanism Theory, Representational Processing Theory and Executive Function theory will not be supported in the current work. First, given the lack of empirical support for the Theory of Mind Mechanism Theory in previous work (see, Leekam et al., 2008; Liu et al., 2018; Sabbagh, Moses & Shiverick, 2006), it is expected that the current study will also not support for this domain-specific account of false belief reasoning.

Second, it is predicted that performance on false signs and metalinguistic tasks will differ because the content of the representations (physical versus propositional) of these tasks is different. This finding would be inconsistent with the Representational Processing Theory, which suggests that both mental and non-mental representational understanding develop via a domain-general change in metarepresentational understanding. Third, it is predicted that the Executive Function theory will also not be supported because previous research suggests there is something unique about false belief understanding over and above conflict inhibition (Liu, Wellman, Tardif & Sabbagh, 2008).

I predict that the Representational Content Theory will be supported. To review, the Representational Content Theory argues that propositional content of the representation is particularly important to reasoning metarepresentational development in mental and non-mental domains. In accordance with the Representational Content Theory, it is predicted that a model predicting False belief task performance from Synonym and Homonym Tasks while controlling for conflict inhibition, expressive vocabulary and age will be the best fitting model (Model 2). Specifically, Model 2 will have a superior fit than Model 3 which predicts false belief task performance from false sign task performance, controlling for conflict inhibition, language and age. Additionally, it is predicted that the current model will replicate age-related advancements on the false belief, false sign and metalinguistic tasks. For a review of the key processes involved in false belief understanding as predicted by each theory, see Figure 6.

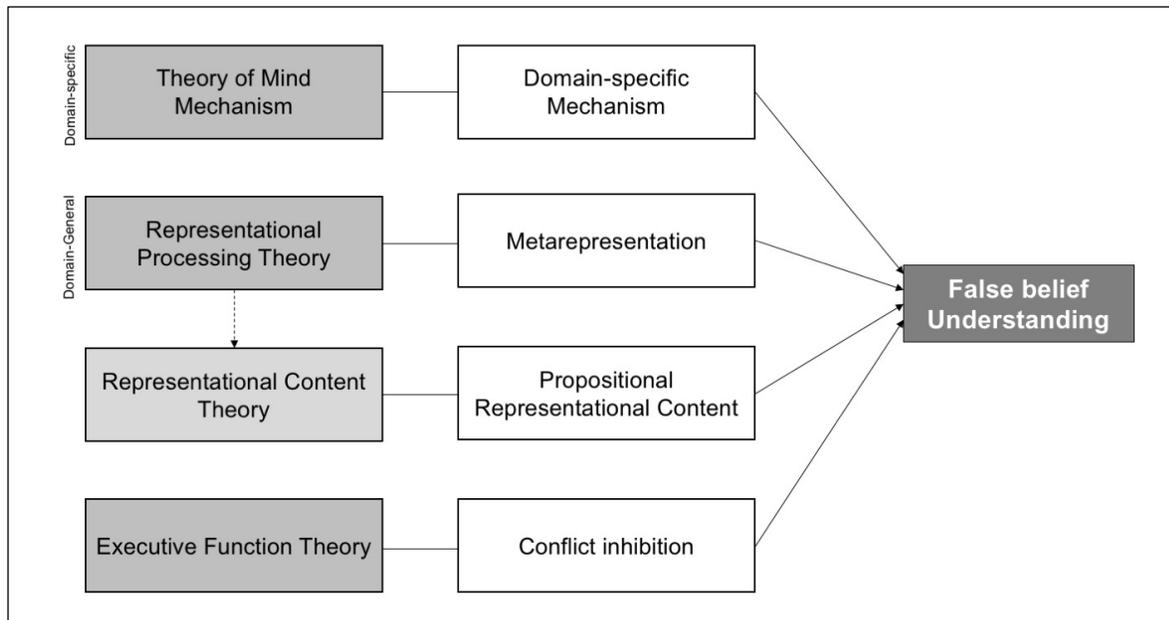


Figure 6. Predictions regarding the key process involved in false belief understanding by theory.

The specific hypotheses of the current work are as follows:

- (1) Metalinguistic task performance will not predict false belief task performance, partialling out the effects of metalinguistic awareness, representation (false signs), chronological age, language abilities, and conflict inhibition (Model 1; Theory of Mind Mechanism Theory).
- (2) Metalinguistic task performance will significantly predict false belief understanding, partialling out the effects of chronological age, language abilities, and conflict inhibition (Model 2).
- (3) Model 2 will have a better fit than a model predicting false belief task performance using false sign tasks as predictors, partialling out the effects of chronological age, language abilities, and conflict inhibition (Model 3).

- (4) Model 2 will have a better fit than a model predicting false belief task performance using conflict inhibition as a predictor, partialling out the effects of chronological age and language abilities (Model 4).
- (5) Children will perform better of the Colour-Colour task than the both homonym and synonym task, respectively.
- (6) Age-related changes will be evident on children's performance on false belief tasks, false sign tasks, and metalinguistic tasks.

Methods

Participants

One-hundred and four (M age = 47.87 months, SD = 4.92 months, range = 38.72 – 56.77 months; 41 males; 63 females) three (n = 48) and four (n = 56) year old children participated in this study. Participants were a convenience sample recruited through the Child Development Lab at the University of Victoria and from preschools in Victoria. Fifty-two (50%) were collected through the Child Development Lab, and fifty-two (50%) were collected through preschools. Children with developmental disorders were excluded from the study. An additional five children participated in the current study but were excluded. Two children were excluded because they refused to complete more than 50% of the tasks. Additionally, one twin from each of the three sets (n = 3) was randomly selected to be excluded in order to maintain the independence of the data. All children spoke English as their first language, and 2.86% (n = 3) children identified as bilingual (2 = English/French; 1 = English/Chinese).

Measures

False belief tasks.

Change in location task. The change in location false belief task (adapted from Wellman & Liu, 2004) was used to evaluate children's explicit false-belief understanding. In this task, children were introduced to a wooden figurine called Scott, who had lost his mittens. Then, a laminated picture of a backpack and closet were placed on the table. Children were told that Scott's mittens are really in his backpack, but that Scott thinks his mittens are in the closet. Then children were asked where Scott will look for his mittens (target question) and where Scott's mittens are really (control question). The task was scored out of 1 (0 = incorrect; 1 = correct). To

receive a score of 1, children were required to answer both the target question correctly (i.e., “closet”) and control question correctly (i.e., “backpack”) (see Appendix A for task scripts).

Unexpected contents task. The unexpected contents task was also used as a measure of false belief understanding. In this task, adapted from Gopnik and Astington (1989), children were shown a Smarties candy box, and asked what they thought was inside (e.g., Smarties). The box was then opened to reveal that there were actually pencils inside, and closed back up. Next, a new character, Penguin Pete (a stuffed animal), was introduced, and children were told that Penguin Pete had never seen inside the box. Children were then asked what Penguin Pete will think is in the box (target question), and what is in the box really (control question). This task was scored out of 1 (0 = incorrect; 1 = correct). To receive a score of 1, children had to answer both the target question correctly (i.e., “Smarties”) and the control question (i.e., “pencils”) correctly (see Appendix B for task scripts).

Metalinguistic tasks.

Synonym Judgement task. The Synonym Judgement Task (Doherty & Perner, 1998) was used as a measure of children’s metalinguistic awareness, specifically their understanding that two different words can refer to the same object. This task consisted of three phases: a vocabulary check phase, a modelling phase, and a test phase. In the vocabulary check phase, children were presented with A4 sized laminated paper that contained four images (two target items, and two distractors). Children were then asked to point to a variety of pictures to ensure that they understood the meaning of the synonym pairs. For example, children were asked to point to the ‘woman’, then, in a subsequent trial, children were asked to point to the ‘lady’ (demonstrating an understanding of the synonym pair). This process was repeated for all

synonym pairs. In order to proceed to the modelling and test phases, children had to correctly identify at least 4 out of the 6 synonym pairs.

In the modelling phase, children were introduced to a puppet, Goofy, who was going to play the game with them. Children were asked to provide one name for an object indicated by the experimenter (e.g., bunny). Children were then told it was the Goofy's job to say the other name (e.g., rabbit), not the name they said. After being reminded of the name they had chosen for the object, children were asked what Goofy should say. Children responded to the question, and then the experimenter provided (or repeated) the correct response, and explained why it would be incorrect for Goofy to say the same name as the child (e.g., bunny), or an alternative name (e.g., house).

In the test phase, children were reminded how to play the game, and then asked to name an object (e.g., mug). After their response, Goofy provided either a correct (e.g., cup) or an incorrect (e.g., star) response, and children were asked if that was what Goofy was supposed to say (test question). Children were given a score of 1 if they correctly answered the test question, and a score 0 if they did not. Three test trials were administered, therefore the task was scored out of a total of three. See Appendix C for scripts.

Homonym Judgement task. The 'Homonym Judgment task' (Doherty, 2000) followed the same procedure as the Synonym task (vocabulary, model, and test phases); however, in this task children were asked to identify two different objects that have the same name. All pictures were displayed on an A4 page which had two target items (i.e., knight/night) and two distractors (i.e., bus, monkey). In the vocabulary check phase, children were asked to point to a variety of items to ensure they understood the meaning of all homonym pairs. To continue to the model and test phase children had to correctly identify at least 4 of the 5 homonym pairs.

In the modelling phases, children were introduced to a puppet, Donald Duck, who was going to play the game with them. Children were asked to provide one name for the object indicated by the experimenter (e.g., bat). Children were then told it was Donald Duck's job to point to the other picture with the same name. After being reminded which object they selected (e.g., baseball bat), children were asked which picture Donald Duck should point to. The child pointed to an item, and then the experimenter responded by confirming or correcting the child's response by pointing correct response (e.g., flying bat). This was followed by an explanation as to why it would be incorrect for Donald Duck to point to the same picture the child pointed to or an item with a different name.

In the test phase, children were reminded of how to play the game, and then asked to point to a named object (e.g., finger nail). Next, Donald Duck either pointed to the correct image (e.g., hardware nail) or an incorrect image (e.g., monkey). After each response, children were asked if that is what Donald Duck should've done (test question). Children were given a score of 1 if they correctly answered the test question, and a score 0 if they did not. Three test trials were administered, therefore the task was scored out of a total of three. See Appendix D for scripts.

Control measures.

Conflict inhibition.

Grass-snow task (Carlson & Moses, 2001). In this task, two cut out felt hand prints were placed on the table in front of the child, and a 15 cm by 15 cm green card and white card were placed to the child's left and right, respectively. To begin, children were asked what colour snow was and what colour grass was to ensure they understood this association. Children were then instructed to place their hands on the cut-out hand shapes, and were told that they were going to play a silly opposites game. The experimenter explained that in this game children were to point

to the white card when they said ‘grass’, and the green card for ‘snow’, and return their hands back onto the felt after each point. Two practice trials were administered. If children did not respond correctly, the experimenter provided corrections. If after three corrections children did not understand the game, the game was over and the child received a score of 0. Children who passed the practice trial completed 16 consecutive test trials. Children received 2 points if they responded correctly, 1 point for a self-corrected response (i.e., pointed to the incorrect card originally and then corrected their response), and 0 if they responded incorrectly. The task was scored out of a total of 32. See Appendix E for scripts.

Head-Toes-Knees-Shoulders task. The Head-Toes-Knees-Shoulders task (McClelland et al., 2014) was designed to evaluate children’s ability to inhibit a typical response in favour of an alternative response. To begin, children were asked what they would normally do if they were asked to touch their head (e.g., touch their head), or touch their toes. Children were then told that they were going to play a silly opposites game. In the silly, game when the experimenter asked them to touch their head they should touch their toes (and vice versa ‘touch your toes’ they should touch their head). Children completed four practice tests and then continued to play the ‘opposites game’ for 10 more trials. Children received a score of 2 for a correct response, a score of 1 for a self-corrected response (i.e., responded incorrectly and self-corrected to the correct action), and a score of 0 for an incorrect response. Children had to receive a score of 10 or greater (out of a total of 20) to continue to the second part of the task. In the second part, two additional commands were added: touch your knees and touch your shoulders. Children were reminded that we were playing the silly game (e.g., if the experimenter says touch your knees they should touch their shoulders, and vice versa) and then completed four practice trials with the new commands. After the practice had been completed, all four actions were combined (i.e.,

head, toes, knees, shoulders) for 10 additional test trials. Trials were scored using the same scoring method as described previously (2 – correct; 1- self-correct; 0 - incorrect) and children were given a score out of twenty. Children were then given an overall score out of forty (combining both the head-shoulders, and head-shoulder-knees-toes phases). The commands were all given in the same fixed order. For script, see Appendix F.

False Sign Tasks (Non-Mental Representations).

False Sign (location). The false sign location task (Sabbagh, Moses & Shiverick, 2006) was used to measure children's understanding of non-mental representations. In the first part of the task, children were introduced to a character, Kim (3-inch wooden figurine), and shown a board with an arrow on a cardboard card that either pointed to the red house or the green house (see Appendix F for photograph of materials). Children were then told that Kim uses the arrow to let people know where she is playing. The arrow was then moved from house to house (four times) and each time the arrow was moved children were asked where the arrow showed Kim was playing. These trials were completed to ensure children understood the meaning of the arrow. Children were then introduced to a new character, Sam (3-inch wooden figurine), and were told that Sam and Kim were trying to decide which house to play in. As the narrative proceeded, children were told that Sam is hungry and needs leaves to get a snack, but before he went he asked Kim how he will know how to find her when he gets back. Kim reassured him that she will leave the arrow pointing to which house she is playing in so she can find him. Sam was then removed from the table, and the arrow was moved to the red house (and Kim was put inside the red house). Children were then told that Kim was going to play in the green house instead (and she is moved to that location), but Kim had forgotten to change the arrow. Then, the experimenter brought Sam back to the table, and asked the children where the arrow said Kim

was (representational-change question), and where Kim was really (control question). Children received a score of 1 if they answered both the representational change question (e.g., red house) and the control question (e.g., green house) correctly, and a score 0 if one or both questions were answered incorrectly. See Appendix G for scripts.

False Sign (contents). The false sign task (adapted from Ioa, Leekman, Perner, McConachie, 2011) tests children's understanding of the representational quality of labels. In this task, children were shown a crayon box and asked what they thought was inside. Children were then shown that there were actually candles inside, and the box was closed. Children were then asked what they now thought was in the box (control question), and what the label shows is inside the box (test question). Children received a score of 1 if they answered both the control question (i.e., candles), and the test question (i.e., crayons) correctly, and a score of 0 if one or both questions were answered incorrectly. See Appendix H for scripts.

Alternative-Naming Task.

Colour-colour task. The Colour-Colour test, adapted from Perner, Stummer, Sprung and Doherty's (2002) alternative naming task, is designed to assess children's ability to name one of two colours displayed on an object. The Colour-Colour task has three phases: the vocabulary check phase, the model phase and the test phase. In the vocabulary check phase, children were shown two A4 sheets each with 4 colours and were asked to point to the named colours. If children were able to correctly identify all 8 colours they proceeded to the modeling phase. In the modeling phase, children were introduced to a puppet, Mickey Mouse, who was going to play the game with them. Children were then shown a 20 cm by 20 cm paper with an image containing two colours (i.e., a green and yellow tree). Then children were told that it was the puppet's job to say one colour (e.g., green) and it was their job to say the other colour (e.g.,

yellow). After children responded, the experimenter provided (or repeated) the correct response (e.g., green) and explained why this was the correct choice, and also indicated why saying the same colour as Mickey (e.g., yellow) or an alternative colour (e.g., pink) would be incorrect. In the test phase, children were first reminded of the rules of the game. Then, children were shown three new two-coloured images in succession. With each image, children were asked to name one colour on the picture, then Mickey would name a colour (providing either a correct or incorrect response), and children were asked if that was what Mickey should have said. Children received a score of 1 for a correct response, and a score of 0 for an incorrect response for each test trial (out of a total of three). See Appendix I for task script.

Expressive Vocabulary subtest from the Kaufman Assessment Battery for Children, 2nd edition.

The Kaufmann Assessment Battery for Children, second edition (K-ABC-II) Expressive vocabulary sub-test was used to evaluate children's general expressive language abilities. In this task, children were shown a standardized series of images, and asked to name each picture. Children received a score of 1 for a correct response, and a score of 0 for an incorrect response for each trial. The task was discontinued after four consecutive scores of zero. Scores were calculated by subtracting the number of errors from the ceiling item.

Procedure

Children were tested individually by an experimenter in one 30 minute session either at Child Development Lab at the University of Victoria or their preschools. The ethics board at the University of Victoria approved the current work. Written consent was obtained from the preschool supervisor (where applicable) and the child's parent or guardian prior to testing. Additionally, verbal assent was obtained from each child before beginning the tasks. In both the

Child Development Lab and in Preschool setting, children were tested in quiet space with minimal distractions. A fixed task order was selected in order to separate tasks which made similar cognitive demands and to facilitate comparison among tasks (see, Carlson & Moses, 2001). Tasks were administered in the following fixed order: K-ABC, false belief – unexpected contents task, synonym judgement task, false sign (location) task, head-shoulders-knees toes task, colour-colour task, snow/grass, homonym judgement task, false belief – location-change task, and the false sign (contents) task.

Child Development Lab.

Children tested in the Child Development Lab were part of a larger longitudinal study. Tasks were presented to the children in the same fixed order at the beginning of a longer testing session. Therefore, fatigue and interference of other cognitive measures were not a concern. Children received stickers throughout their participation. Additionally, to thank families for participating in the longitudinal project, children received a small prize after completing the tasks, and parents received a \$20 honorarium.

Preschools.

Children tested at preschools also completed the tasks in the fixed order described above. Children received stickers throughout their participation. The preschool centre received a picture book as a token of appreciation for their participation.

Data Analysis Plan

The hypotheses of the current work was tested using both a traditional null hypothesis statistical testing (NHST) approach, as well as a Bayesian hypothesis testing approach. In the NHST approach, zero-order correlations was presented for all measures, and aggregate scores. Partial correlations will be estimated for false belief, false sign, and metalinguistic tasks

(controlling for conflict inhibition and expressive vocabulary). Gender differences in performance was evaluated using independent samples *t*-tests. Hierarchical linear regressions was used evaluate the proposed models (See Table 1). Simple linear regression was used to predict task performance (false belief, false sign, and metalinguistic) from age (in months). Finally, Chi-Square Tests was used to compare performance on the alternative naming tasks (the Synonym Judgement task, Homonym judgement task and the colour-colour task).

All data was also analysed using Bayesian analysis methods in order to provide a more comprehensive and informative statistical assessment of both the null and alternative hypothesis (Ly et al., 2017). The Bayesian approach was used in addition to the NHST approach because it provides information on the likelihood of the data given the null and alternative hypothesis. The use of the Bayesian approach is especially beneficial in the current work because this method allows us to determine if the data is inconclusive (i.e., no evidence for either the null or alternative hypothesis). In other words, if NHST approach suggest we should accept the null hypothesis, the Bayesian model can determine if the null is more likely than the alternative, or if more evidence is required (i.e., the data is inconclusive). Bayesian correlations was presented for each measure to determine the size of the effect of the associations. Additionally, Partial Bayesian correlations was used to determine the amount of evidence for the association between false belief tasks, false sign tasks, and metalinguistic tasks (controlling for conflict inhibition and expressive vocabulary). Gender differences was evaluated using Bayesian Independent *t*-tests. Bayesian linear regression analyses was used to predict performance on the false belief, false sign, metalinguistic, conflict inhibition and language tasks from age (in months). Bayesian linear regression analyses was conducted to evaluate the likelihood of each of the proposed models in

comparison to a null model, as well as in comparison to the best possible model. Finally, Bayesian contingency tables was used to compare performance on the alternative naming tasks.

In the NHST framework, scientists draw inferences from their data in such a way that over the long run (i.e., many replications from random samples of the same population), they are able to maintain a particular (and low) error rate for their conclusions. In most psychological research, scientists are concerned with the possibility of making Type I errors (i.e., false positives), and therefore draw inferences only when the long-run possibility of their making a Type I error is small (e.g., 5%). A *p*-value is the probability of observing a given effect (or one more extreme) given that the null hypothesis is true, and if a scientist rejects the null hypothesis when $p < .05$ (when the rules of NHST are followed), they should only be mistaken about the null hypothesis approximately 5% of the time.

The *p*-value, however, does not quantify the likelihood of the null hypothesis (H_0) or the alternative hypothesis (H_1). A *p*-value $< .05$ is often interpreted as strongly in favour of the alternative hypothesis over the null hypotheses; however, NHST analyses do not quantify the support for the null hypotheses nor the alternative hypotheses. Therefore, accepting an alternative hypothesis via a traditional NHST method is logically invalid (Wagenmakers et al., 2018a). As a result of the limitations of the information provided by the *p*-value in NHST approach, Bayesian analyses will also be conducted to supplement the results.

Bayesian analyses will extend the classical NHST statistical testing approach by providing a more comprehensive assessment of the hypotheses. Bayesian hypothesis testing allows for the quantification of evidence for each hypothesis (Ly et al., 2017). Bayesian analyses are based on a probability theory which can be used to test the likelihood of a given hypothesis via the bayes factor (BF). A BF is defined as: “the degree to which the data sway our belief from

one to the other hypothesis” (Etz & Vandekerckhove, 2016, p.4). BFs provide evidence (1) in favour of the alternative hypothesis (H_1 ; evidence in favour of an effect) (2) in favour of the null hypotheses (H_0 ; evidence in favour of an absence of an effect); and (3) evidence that is ambiguous, favouring neither H_1 or H_0 (Wagenmakers et al., 2018b). The interpretation of the BF changes via the numerator and denominator in the ratio. The BF_{01} is the likelihood of the null hypotheses over the alternative hypothesis, and is interpreted as evidence in favour of the null hypotheses. For example, a BF_{10} ($1/BF_{01}$) is H_1 over H_0 , and is interpreted as the likelihood of the outcome under the alternative hypothesis. For example, a $BF_{10} = 15$ means that the observed data are 15 times more under H_1 than H_0 (no effect). Whereas, the inverse, a $BF_{01} = 4$ would indicate that the data are 4 time more likely under H_0 than under H_1 . Additionally, Bayesian modeling approach allows for the specification of the prior distribution. The Bayesian hypothesis testing approach allows for quantification of the data in support for H_1 relative to H_0 .

Data Strategy

The central advantage of the Bayesian analyses, in comparison to an null hypothesis significance testing (NHST) approach, is the ability to quantify evidence in favour of H_1 in comparison to H_0 . By contrast, the NHST approach uses the p -value to determine significance, which is conditional on the null hypothesis being true, In other words, the BF allows for quantification of the amount of evidence (the degree of belief) in favour of each hypothesis, rather than simply detecting whether the observed data are (un)likely to have come from a population where the null hypothesis is true. The BF is a “thermometer for the intensity of the evidence”, and provides the likelihood of each hypotheses being true (Wagenmakers et al., 2018a, p. 46). In the Bayesian model, hypotheses are not “violently biased against the null” (like

the classical significance tests) but rather evaluate the likelihood of evidence in favour of the null given the data.

Additionally, data collection does not need to reach a specific set-point before Bayesian analyses can be completed. This is especially helpful for studies with difficult to recruit populations, such as young children. Data collection can continue until the evidence is compelling, or resources have diminished and hypotheses can be evaluated as data is accumulated (Wagenmakers et al., 2018a). Although, Bayes Factors are limited by small sample sizes, the results will indicate that the data is inconclusive and no conclusions can be drawn in favour of the null or alternative hypothesis. Whereas, in frequentist approaches, small sample sizes often result in greatly underpowered studies, which is especially common in psychological science (Wagenmakers, Verhagen & Ly, 2016). For a review of the advantages of Bayesian inference over classical NHST, see Wagenmakers and colleagues (2018a).

Interpreting bayes factors.

The Bayes Factor (BF) provides a continuous scale of evidence for H_1 relative to H_0 , which can be interpreted using cut-offs provided by Jeffreys (see Table 11). When using the Bayesian inference approach with BFs, there is a potential for false positive evidence (FPE; strong evidence for H_1 when H_0 is correct) or false negative evidence (FNE; strong evidence for evidence H_0 when H_1 is correct). To avoid misleading evidence, evidential thresholds for BF, or design priors, must be set a priori and should be based on the expected research design, and sample size. The current analyses uses a 'fixed n' design. Within this design, data collection was stopped based on availability of resources at a sample size ($n = 104$), and the strength of evidence (and misleading evidence) was determined a priori based on this selected design. According to Schönbrodt and Wagenmakers (2018) for a sample of $n = 100$ ($\delta = 0.5$) (using a

fixed n design), BF thresholds of $H_1 = 6$ and $H_0 = 1/6$ are reasonable. These evidential thresholds drop the FPE rate from 0.9% to 0.6% and the dropping the FNE rate from 0.3% to virtually 0%. For the purposes of the current work, $1/6 < BF < 6$ were considered inconclusive (i.e., no sufficient evidence for H_1 or H_0) (based on the simulation from, Schönbrodt & Wagenmakers, 2018). By setting these specific evidential thresholds, the rate of finding misleading evidence (i.e., FPE or FNE) lowers considerably.

Table 2.

*A basic heuristic for interpreting Bayes Factors
(adapted from Jeffreys, 1961).*

Bayes Factor	Interpretation
> 100	Extreme evidence for H_1
30 - 100	Very strong evidence for H_1
10 - 30	Strong evidence for H_1
3 - 10	Moderate evidence for H_1
1 - 3	Anecdotal evidence for H_1
1	No evidence
1/3 - 1	Anecdotal evidence for H_0
1/10 - 1/3	Moderate evidence for H_0
1/30 - 1/10	Strong evidence for H_0
1/100 - 1/30	Very strong evidence for H_0
< 1/100	Extreme evidence for H_0

Informed prior distributions.

A prior probability distribution, often referred to simply as the prior, is the expected belief in the posterior distribution before data is evaluated. In the current work, the prior used for all Bayesian analyses was the JZN prior with a default prior scale. The JZN prior assumes that the effect sizes under H_1 follow a Cauchy distribution. The Cauchy prior width (the default in JASP) is used for the effect size of the model. The Cauchy prior implies when $r=x$ then 50% of the values lie between the positive x-value and the negative x-value. For example, with $r = .85$,

50% of values lie between -0.85, and +0.85. The Cauchy prior is centered around (a median) of zero and was used for all Bayesian analysis.

Results

Data Preparation

Data was analyzed using R (Version 1.1.447) and JASP (Version 0.8.6) [Computer software]. All frequentist analyses—descriptive statistics, *t*-tests, regression analyses, and chi-square tests—were calculated in R using the following packages: ‘car’ (Fox & Weisberg, 2011), ‘psych’ (Reville, 2018), ‘apaTables’ (Stanley, 2018), ‘ppcor’ (Kim, 2015), ‘ggplot2’ (Wickham, 2009). Bayesian analyses were calculated using JASP (<https://jasp-stats.org>) and in R using the ‘BayesMed’ (Nuijten, Wetzels, Matzke, Dolan & Wagenmakers, 2015) and ‘BayesFactor’ packages (Morey & Rouder, 2015).

All data were complete: there were no missing cases on any measure for any participant. Data entry was reviewed to ensure no input errors were made. No outliers outside of 95% confidence interval were identified. Aggregate scores (sums of scores on tasks measuring the same construct) were created prior to analyses. For example, scores on the false belief location task and the false belief contents task were added to produce an aggregate false belief score. Aggregate scores were also created for metalinguistic tasks (Synonym Judgment Task + Homonym Judgment Task), false sign tasks (false sign location + false sign contents), and for conflict inhibition tasks (Grass/Snow + Head-Shoulder-Knees-Toes).

Descriptive Statistics

Table 2 shows descriptive statistics for all individual tasks completed. Table 3 shows descriptive statistics for the aggregate scores for false belief tasks, false sign tasks, metalinguistic tasks, and conflict inhibition tasks. Independent *t*-tests show no significant ($p < .05$ level) gender differences on expressive vocabulary, conflict inhibition, false belief tasks, false sign tasks, or metalinguistic tasks (see Table 4). The percent of control questions answered correctly and the percent of correct responses overall on the false belief, false sign and metalinguistic tasks are

displayed in Table 5. Table 6 shows the percent of vocabulary items known on the Synonym Judgement Task and Homonym Judgement Task, respectively. An average scaled score for overall performance on the K-ABC revealed that, in general, children's expressive vocabulary task performance was in the high average range.

Bayesian independent t-tests with Cauchy prior centred at zero and a standard prior width set to 0.707 were computed in JASP to determine if there were gender differences on expressive vocabulary, conflict inhibition, false belief tasks, false sign tasks, or metalinguistic tasks. The data supports neither the null hypothesis (no gender differences) nor the alternative hypothesis (gender differences) on all dependent measures (expressive vocabulary, conflict inhibition, false belief tasks, false sign tasks and metalinguistics tasks) the Bayes Factors are not sensitive to the prior widths. All BF_{10} fell in the inconclusive range ($1/6 < BF < 6$), therefore no conclusions regarding gender differences can be made at this time. Further data collection is required. See Table 4 for BF_{10} values.

Table 3.

Mean, range and standard deviations of all variables (N = 104).

Variable	N	Range	M	SD
Age (months)	104	38.73 – 56.78	47.87	4.49
False Belief - Contents	104	0 - 1	0.28	0.45
False Belief – Location	104	0 - 1	0.39	0.49
False Sign - Contents	104	0 - 1	0.63	0.49
False Sign - Location	104	0 - 1	0.72	0.45
Synonym Judgment task	104	0 - 3	2.12	1.17
Homonym Judgement task	104	0 - 3	1.34	1.41
Colour-Colour	104	0 - 3	2.67	0.70
Grass/Snow	104	0 - 32	11.55	12.72
Head-Shoulders-Knees-Toes	104	0 - 40	19.15	11.44

Expressive Vocabulary	104	9 - 30	20.03	3.89
-----------------------	-----	--------	-------	------

Note. M and SD are used to represent means and standard deviations, respectively.

Table 4.

Mean, range and standard deviations for aggregate scores (N = 104).

Variable	N	Range	M	SD
False Belief Tasks	104	0 – 2	.66	.76
False Sign Tasks	104	0 – 2	1.35	.76
Metalinguistic Tasks	104	0 – 6	3.45	2.05
Conflict Inhibition Tasks	104	0 – 71	30.70	21.45

Note. M and SD are used to represent means and standard deviations, respectively.

Table 5.

Independent Sample t-tests and Bayesian Independent t-tests predicting task performance from gender.

Dependent Variable	<u>Frequentist</u>		<u>Bayesian</u>	
	<i>t</i> (df)	<i>p</i> -value	BF ₁₀	error %
Expressive vocabulary	.78(102)	.43	0.278	0.035
Conflict inhibition	1.38(102)	.17	0.491	5.208e -5
False belief tasks	-.47(102)	.34	0.234	0.032
False sign tasks	.58(102)	.57	0.245	0.033
Metalinguistic tasks	-1.123(102)	.26	0.370	6.188e -5

**Note.* Gender is the predictor. Frequentist refers to a traditional Student's independent t-test approach. df refers to degrees of freedom. Bayesian refers to the results of the Bayesian Independent samples t-test. BF₁₀ is the evidence in favour of the alternative hypothesis over the null hypothesis. Bayesian analyses use a Cauchy (prior width = 0.707).

Table 6.

Percent of correct responses on control questions and overall.

Task	<u>Percent Correct</u>	
	Control Question	Overall
False Belief Contents	94.23	27.89
False Belief Location	56.62	38.46
False Sign Contents	87.50	62.50
False Sign Location	79.81	72.12
Synonym Judgement Task	—	55.70
Homonym Judgement Task	—	35.70

**Note.* Percent correct refers to the percent of children who passed the False belief tasks (score: 1/1), False Sign tasks (score: 1/1), and received a perfect score on the Synonym and Homonym Judgement Task (score: 3/3, respectively).

Table 7.

Percent of children who identified both items from synonym or homonym pairs in the vocabulary check, and the percent of children who passed the vocabulary check.

Synonym Pairs	% known	Homonym Pairs	% known
Bunny / Rabbit	91.3	Bat (baseball/animal)	58.7
Cup / Mug	76.9	Letter (mail/alphabet)	51.0
T.V. / television	87.5	Nail (finger/hardware)	63.5
Present / gift	81.7	Glasses (spectacles/drinking)	74.0
Street / Road	95.2	Night /Knight	73.1
Woman / Lady	93.3		
Passed (4/6)	95.0	Passed (4/5)	50.9

**Note.* To pass the vocabulary check on the Synonym task children must identify 4 out of 6 synonym pairs. To pass the vocabulary check on the homonym task children must identify 4 out of 5 homonym pairs.

Correlations

Zero-order correlations. Zero-order correlations for all tasks are displayed in Table 7.

All tasks measuring the same construct (i.e., false belief contents and false belief location change) were significantly correlated at the $p < .05$ level. The false belief tasks were not significantly correlated with any other measure. The Grass/Snow (conflict inhibition) task was significantly correlated with all tasks (except the false belief tasks). The Head-Shoulders-Knees-Toes (conflict inhibition) task was significantly correlated with all tasks (except the false belief tasks, and the false sign contents task). Expressive vocabulary was significantly correlated with all tasks (except the false belief tasks and the Homonym Judgement Task). Table 8 shows zero-order correlations for the aggregate scores. The zero-order correlations follow a similar pattern as seen with the correlations among individual tasks. Conflict inhibition was significantly correlated with the metalinguistic and false sign tasks respectively, but not with the false belief tasks. Expressive vocabulary was also significantly correlated with the metalinguistic and false sign tasks, but not the false belief tasks. Overall, correlation analyses indicate the control variables (conflict inhibition and expressive vocabulary) were significantly correlated with false sign and metalinguistic tasks, and false belief tasks were not significantly correlated **with any measure**.

Bayesian correlations. Bayesian correlations with Cauchy prior centred at zero with a standard prior width set to 0.707 were calculated in JASP. In Bayesian correlation analyses, data which is more likely under (or favour) the alternative hypothesis is interpreted as evidence for an association between the constructs; whereas, data which is more likely under (or favour) the null hypothesis is interpreted as evidence for a lack of association between constructs. Table 6 shows the BF for the association between all individual tasks included in the current analyses. BF_{10} factors indicate a similar trend that was identified in the frequentists analyses.

Table 6 also shows the correlation between the control tasks—conflict inhibition and expressive vocabulary—and the core constructs. The $BF_{10} > 6$ criterion was used to evaluate the evidence in favour of the alternative and null hypotheses. The Bayesian correlation analyses revealed the data was more likely under the alternative hypothesis than the null hypothesis for the association between control tasks (conflict inhibition tasks (Grass/Snow and Head-Shoulders-Knees-Toes) and expressive vocabulary) and the metalinguistic tasks and false belief tasks; however, the evidence favoured the null for the association between the control tasks and the false belief tasks.

Table 7 shows the Bayesian correlations between the aggregate scores for each task—conflict inhibition, false belief, false sign, and metalinguistic tasks—and expressive vocabulary. An association between conflict inhibition and metalinguistic tasks were 20663.84 times more likely under the alternative than the null hypothesis. Additionally, an association between conflict inhibition and the false sign task was 21.81 times more likely under the alternative than the null hypothesis. Additionally, these analyses indicated that data favoured the alternative hypothesis for the association between expressive vocabulary and the following tasks: conflict inhibition task performance ($BF_{10} = .36.28$), metalinguistic tasks ($BF_{10} = 38.63$) and false sign tasks ($BF_{10} = 29.29$). However, evidence was 8.33 times more likely under the null hypothesis for the an association between the false belief and metalinguistic tasks, and 7.14 times more likely under the null for an association between false belief and false sign tasks. All other Bayesian correlations did not provide sufficient evidence in favour of either the alternative nor null hypotheses.

Table 8.

Zero-order correlations and Bayesian Correlations between all false belief, false sign, metalinguistic, executive function and control tasks.

Variable	1	2	3	4	5	6	7	8	9	10
1. False Belief - Contents	—									
2. False Belief – Location	.30** (14.56)	—								
3. False Sign - Contents	-.01 (.12)	-.04 (.13)	—							
4. False Sign - Location	.20* (.86)	.01 (.12)	.32** (23.35)	—						
5. Synonym Judgment task	.09 (.18)	-.03 (.13)	.13 (.28)	.17 (.55)	—					
6. Homonym Judgement task	-.04 (.13)	.02 (.13)	.12 (.24)	.09 (.18)	.26** (3.94)	—				
7. Colour-Colour	.08 (.16)	-.03 (.13)	.18 (.61)	.05 (.14)	.24* (2.13)	.14 (.34)	—			
8. Grass/Snow	.07 (.14)	.05 (.14)	.24* (.49)	.28** (1.45)	.34*** (32.42)	.31*** (41.74)	.21* (7.11)	—		
9. Head-Shoulders-Knees-Toes	.05 (.16)	.06 (.14)	.17 (2.15)	.22* (7.13)	.33*** (63.15)	.33*** (19.03)	.28** (1.10)	.57*** (6178e+4)	—	
10. Expressive Vocabulary	.14 (.33)	.14 (.32)	.20* (.87)	.33*** (42.03)	.41*** (1095.71)	.14 (.34)	.28** (7.47)	.29** (10.43)	.29** (10.12)	—

Note.* * indicates $p < .05$; ** indicates $p < .01$; * indicates $p < .001$. Pearson's r correlation value (top) and Bayes factor (BF_{10}) is indicate below in brackets. BF_{10} indicates evidence in favour of the alternative hypothesis.

Table 9.

Zero-order correlations and Bayesian Correlations for aggregate scores.

Variable	1	2	3	4	5
1. Expressive Vocabulary	—				
2. Conflict inhibition	.34*** (36.28)	—			
3. False Belief tasks	.17 (.54)	.08 (.17)	—		
4. Metalinguistic tasks	.33*** (38.63)	.46*** (20663.84)	.01 (.12)	—	
5. False Sign tasks	.32*** (29.29)	.31** (21.81)	.05 (.14)	.19 (.79)	—

Note. * indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$. Pearson's r correlation value (top) and Bayes factor (BF_{10}) is indicate below in brackets. BF_{10} indicates evidence in favour of the alternative hypothesis.

Partial correlations.

As a result of the significant correlations between the control measures (conflict inhibition and expressive vocabulary) and the false sign and metalinguistic tasks, a partial correlation controlling for these variables was computed. See Table 9 for Pearson's r values and significance for the partial correlations partialling out the effects of conflict inhibition and expressive vocabulary, respectively. Results from these analyses show that when partialling out the effects of conflict inhibition, expressive vocabulary remained significantly correlated with the false sign tasks, the Synonym Judgement Task and the Colour-Colour task. Similarly, partialling out the effects of expressive vocabulary, conflict inhibition remained significantly correlated with the false sign tasks, the Synonym judgment task, the homonym judgement task and the Colour-Colour task. The Synonym judgement task and the Homonym judgement task remain significantly correlated partialling out the effects of expressive vocabulary.

A subsequent partial correlation analysis was conducted partialling out the effects of both the control tasks (conflict inhibition and expressive vocabulary)(see Table 10). After partialling out the effects these control tasks, there were no significant correlations between the core constructs.

Partial bayesian correlations. Following from the frequentist analyses, Bayesian partial correlations were also conducted to determine the likelihood of the association between constructs partialling out the effects of expressive vocabulary, and conflict inhibition, respectively (See Table 8). Partialling out conflict inhibition, Bayes factor estimates indicated that an association between expressive vocabulary and the synonym judgment task was 59.95 times more likely under the alternative than the null hypothesis (no association); however, evidence favoured the null (7.14:1) for the association between expressive vocabulary and the

homonym judgment task. The association between the false sign task and the homonym judgement task was 7.69 times more likely under the null hypothesis (no association) than under the alternative. All other partial correlations (partialling out conflict inhibition) did not provide sufficient evidence in favour of the alternative nor null hypothesis. Partialling out expressive vocabulary, evidence favoured the alternative hypothesis for the association between conflict inhibition and the Synonym Judgment task ($BF_{10} = 13.81$), and for the association between conflict inhibition and the Homonym Judgment task ($BF_{10} = 96.53$). All other Bayesian correlations (partialling out expressive vocabulary) were inconclusive.

Table 9 shows the Bayesian partial correlations amongst constructs controlling for the combined effects of expressive vocabulary and conflict inhibition. After partialling out the effects of these control tasks, the association between the Homonym Judgement Task and False sign task was 10 times more likely under the null than the alternative hypothesis. All other partial correlations were inconclusive ($1/6 < BF < 6$), therefore no evidence for the alternative nor null hypothesis can be derived given the current data.

Table 10.

Partial correlations and partial Bayesian correlations between all false belief, false sign, and metalinguistic tasks, partialling out conflict inhibition and expressive vocabulary, respectively.

Variable	1	2	3	4	5	6	7
1. False belief tasks	—	<i>-.03(.12)</i>	<i>-.04(.09)</i>	<i>-.04(.09)</i>	<i>-.02(.09)</i>	—	<i>.02(.15)</i>
2. False Sign tasks	.03(.15)	—	<i>.06(.20)</i>	<i>.09(.27)</i>	<i>.06(.19)</i>	—	<i>.23*(3.69)</i>
3. Synonym Judgment task	.00(.12)	<i>.08 (.24)</i>	—	<i>.23*(2.79)</i>	<i>.14(.53)</i>	—	<i>.28**(13.81)</i>
4. Homonym Judgement task	<i>-.04(.09)</i>	<i>.01(.13)</i>	<i>.14(.23)</i>	—	<i>.11(.30)</i>	—	<i>.33***(.96.53)</i>
5. Colour-Colour	.00(.13)	<i>.06(.20)</i>	<i>.15(.65)</i>	<i>.05(.17)</i>	—	—	<i>.20*(1.66)</i>
6. Expressive Vocabulary	.15(.43)	<i>.24*(4.99)</i>	<i>.32**(59.95)</i>	<i>.03(.14)</i>	<i>.21*(2.20)</i>	—	
7. Conflict inhibition	—	—	—	—	—	—	—

Note.* * indicates $p < .05$; ** indicates $p < .01$; * indicates $p < .001$. Non-italicized indicates the partial correlations with conflict inhibition partialled out. Italicized indicates partial correlations with expressive vocabulary partialled out. Pearson's r correlation value are presented and Bayes factors (BF_{10}) are presented in brackets. BF_{10} indicates evidence in favour of the alternative hypothesis.

Table 11.

Partial correlations and partial Bayesian Correlations between all false belief, false sign, metalinguistic tasks, partialling out conflict inhibition and express vocabulary.

Variable	1	2	3	4
1. False belief tasks	—			
2. False sign tasks	.01(.13)	—		
3. Synonym Judgment task	-.01(.11)	.05(.17)	—	
4. Homonym Judgement task	-.02(.08)	-.01(.10)	.13(.50)	—
5. Colour-Colour	-.00(.12)	.05(.17)	.14(.57)	.04(.16)

Note.* * indicates $p < .05$; ** indicates $p < .01$; * indicates $p < .001$. Pearson's r correlation value (top) and Bayes factor (BF_{10}) is indicated below in brackets. BF_{10} indicates evidence in favour of the alternative hypothesis.

Regression Analyses

Hierarchical linear regression. As a result of the lack of significant correlations between the false belief tasks (dependent variable) and predictors (metalinguistics tasks, false sign tasks, and conflict inhibition) hierarchical linear regression to test proposed models 1 through model 4 could not be meaningfully conducted.

Bayesian regression analyses. Bayesian Regression analyses were performed in JASP (Love et al., 2015; Morey & Rouder, 2015; Rouder et al., 2012) using a JZS prior scale (r scale = 0.354). The first model compared the alternative model (an effect) with Synonym/Homonym Judgement Tasks, False sign task performance, and conflict inhibition predicting false belief task performance to the null model (which included expressive vocabulary and age in months). The Bayes factors estimates indicate there is insufficient data to predict evidence in favour of the alternative or null hypotheses for an effect of individual predictors—metalinguistic tasks, conflict inhibition tasks, false sign tasks— independently on false belief task performance. An estimated Bayes Factor suggests that a model predicting false belief task performance from false sign and metalinguistic task performance was 6.49 ($BF_{01} = 1/0.154$) times more likely under the

null (no effect) than the alternative. Additionally, an estimated Bayes Factor suggests that a model predicting false belief task performance from false sign tasks and conflict inhibition tasks were 6.85 ($BF_{01} = 1/0.146$) more likely under the null than the alternative hypothesis. Finally, a model predicting false belief task performance from false sign tasks, conflict inhibition, and metalinguistic task performance was 13.51 ($BF_{01} = 1/0.074$) times more likely under the null than the alternative hypothesis. Thus, Bayesian regression analyses suggest the combination of the central predictors in the proposed model was generally more likely under the null (no effect) than the alternative model. See Table 12 for complete model statistics.

A subsequent Bayesian regression analysis predicting false belief task performance was computed adding conflict inhibition to the null model (in addition to age (in months) and expressive vocabulary). After adding conflict inhibition to the null model, the estimated Bayes Factors indicated that the current data were insufficient to find meaningful evidence in favour of the null or alternative models (see Table 13).

Table 12.

Bayesian linear regression model predicting false belief task performance (null model comparison - including age and expressive vocabulary).

Models	P(M)	P(M data)	BF _M	BF ₁₀	R ²
Null model (incl. age, expressive vocabulary)	0.125	0.378	4.251	1.000	0.033
Metalinguistic Tasks	0.125	0.145	1.184	0.383	0.034
Conflict inhibition tasks	0.125	0.138	1.123	0.366	0.033
False sign tasks	0.125	0.133	1.076	0.353	0.033
Conflict inhibition tasks + Metalinguistic Tasks	0.125	0.065	0.485	0.171	0.037
False sign tasks + Metalinguistic Tasks	0.125	0.058	0.431	0.154	0.035
Conflict inhibition tasks + False sign tasks	0.125	0.055	0.408	0.146	0.033
Conflict inhibition tasks + False sign tasks + Metalinguistic Tasks	0.125	0.028	0.203	0.074	0.037

**Note.* All models include age, expressive vocabulary. P(M) is the prior probability of the model. P(M|data) is the posterior probability of the model based on the data. BF_m is the likelihood odds (Bayes Factor) of the model against all other models. BF₁₀ is the likelihood odds of under H₁ over H₀.

An additional Bayesian regression model comparison was computed predicting performance on the false belief task from all predictors using the best possible model as the comparison group. In this model, all potential predictors (including control variables: age, expressive vocabulary and conflict inhibition) were included in the analysis. The strongest evidence is found in favour of the null hypothesis was for an effect of age, metalinguistic tasks, false sign tasks, and conflict inhibition on false belief task performance. Bayes factor analysis estimates that this model was 70.525 ($BF_{01} = 1/0.014$) times more likely under the null hypothesis (no effect) than the alternative hypothesis. Adding expressive vocabulary as a predictor reduced the evidence in favour of the null from 70.525 to 49.731 ($BF_{01} = 1/0.02$) times more likely. Overall, evidence for all combinations of variables favoured the null hypotheses (no effect) over the alternative hypothesis. Additional data is required to determine whether the alternative or the null hypothesis is more likely in these cases. See Table 14 for complete model statistics.

Table 13.

Bayesian linear regression model predicting false belief task performance (null model comparison – including age, expressive vocabulary and conflict inhibition).

Models	P(M)	P(M data)	BF_M	BF₁₀	R²
Null model (incl. age, expressive vocabulary, conflict inhibition)	0.250	0.483	2.803	1.000	0.033
Metalinguistic Tasks	0.250	0.226	0.877	0.468	0.037
False sign tasks	0.250	0.193	0.715	0.399	0.033
False sign tasks + Metalinguistic Tasks	0.250	0.098	0.327	0.203	0.037

Note. All models include age, expressive vocabulary, conflict inhibition. P(M) is the prior probability of the model. P(M|data) is the posterior probability of the model based on the data. BF_m is the likelihood odds (Bayes Factor) of the model against all other models. BF₁₀ indicates evidence in favour of the alternative hypothesis.

Table 14.

Bayesian linear regression model predicting false belief task performance best possible model comparison.

Models	P(M)	P(M data)	BF_M	BF₁₀	R²
Null model	0.031	0.196	7.537	1.000	0.000
EV	0.031	0.159	5.857	0.812	0.029
Conflict inhibition	0.031	0.054	1.767	0.276	0.006
EV + age	0.031	0.053	1.730	0.270	0.033
EV + MLA	0.031	0.051	1.661	0.260	0.032
EV + Conflict inhibition	0.031	0.047	1.532	0.241	0.030
False Signs	0.031	0.046	1.491	0.235	0.003
EV + False Signs	0.031	0.046	1.490	0.234	0.029
MLA	0.031	0.041	1.316	0.208	0.000
Age	0.031	0.041	1.309	0.207	0.000
EV + Age + MLA	0.031	0.020	0.640	0.103	0.034
EV + Conflict inhibition + MLA	0.031	0.020	0.623	0.101	0.034
EV + Age + Conflict inhibition	0.031	0.019	0.611	0.099	0.033
EV + Age + False Signs	0.031	0.019	0.589	0.095	0.033
EV + False Signs + MLA	0.031	0.018	0.564	0.091	0.032
Conflict inhibition + False Signs	0.031	0.017	0.535	0.087	0.007
Conflict inhibition + MLA	0.031	0.017	0.534	0.087	0.007
EV + Conflict inhibition + False Signs	0.031	0.017	0.525	0.085	0.030
Age + Conflict inhibition	0.031	0.017	0.525	0.085	0.007
Age + False Signs	0.031	0.014	0.449	0.073	0.003
False Signs + MLA	0.031	0.014	0.441	0.072	0.003
Age + MLA	0.031	0.013	0.393	0.064	0.000
EV + Age + Conflict inhibition + MLA	0.031	0.009	0.283	0.046	0.037
EV + Age + False Signs + MLA	0.031	0.008	0.254	0.041	0.035
EV + Conflict inhibition + fsu + MLA	0.031	0.008	0.246	0.040	0.034
EV + Age + Conflict inhibition + False Signs	0.031	0.008	0.241	0.039	0.033
Conflict inhibition + False Signs + MLA	0.031	0.006	0.202	0.033	0.008
Age + Conflict inhibition + False Signs	0.031	0.006	0.202	0.033	0.008
Age + Conflict inhibition + MLA	0.031	0.006	0.197	0.032	0.007
Age + False Signs + MLA	0.031	0.005	0.165	0.027	0.003
EV + Age + Conflict inhibition + False Signs + MLA	0.031	0.004	0.122	0.020	0.037
Age + Conflict inhibition + False Signs + MLA	0.031	0.003	0.086	0.014	0.009

**Note.* MLA refers to metalinguistics. EV refers to expressive vocabulary. P(M) is the prior probability of the model. P(M|data) is the posterior probability of the model based on the data. BF_m is the likelihood odds (Bayes Factor) of the model against all other models. BF₁₀ indicates evidence in favour of the alternative hypothesis.

Age-related effects.

Linear regression. Simple linear regression models were computed to determine if age (in months) predicted performance on the key predictors: expressive vocabulary, false belief task performance, false sign task performance and false sign task performance. Age was a significant predictor of expressive vocabulary, conflict inhibition and false sign task performance, with older age predicting better performance. Age was not a significant predictor of false belief task performance, or metalinguistic task performance. See Table 10 for complete model statistics.

Table 15.

Simple linear regressions predicting task performance from age (in months).

Dependent Variable	R ²	Adjusted R ²	df	F	<i>p</i>
Expressive vocabulary	.08	.08	(1, 102)	9.41	< .01
Executive Function	.04	.03	(1, 102)	4.28	< .05
False belief tasks	.000	-.001	(1, 102)	.001	.97
False sign tasks	.09	.08	(1, 102)	10.22	< .01
Metalinguistic tasks	.02	.01	(1, 102)	2.29	.13

Note. Predictor is age in months.

Bayesian linear regression. A Bayesian Linear regression was performed to look at the effect of age (in months) on task performance using the default JZN prior (r scale = 0.354). The analyses revealed that alternative hypothesis that age had an effect on expressive vocabulary performance was 12.34 times more likely than the null hypothesis of no effect. Additionally, the alternative hypothesis that age had an effect on false sign task performance was 17.38 times more likely than the null hypothesis of no effect. The effect of age and all other tasks (conflict inhibition, false belief and metalinguistics) did not revealed a Bayes factor in favour of either the null of the alternative. See Table 15 for complete model statistics.

Table 16.

Bayesian linear regressions predicting task performance from age (in months).

Dependent Variable	P(M)	P(M data)	BF _M	BF ₁₀	R ²
Expressive vocabulary	.50	.96	12.34	12.34	.08
Executive Function	.50	.58	1.37	1.37	.04
False belief tasks	.50	.17	.21	.21	.00
False sign tasks	.50	.95	17.38	17.38	.09
Metalinguistic tasks	.50	.36	.57	.57	.02

Note. Predictor is age in months. P(M) is the prior probability of the model. P(M|data) is the posterior probability of the model based on the data. BF_m is the likelihood odds (Bayes Factor) of the model against all other models. BF₁₀ is the likelihood odds of under H₁ over H₀.

Performance on alternative naming tasks.

Chi-square. A Chi-Square test indicates that there is a significant difference in performance on the Homonym Judgement task and the Colour-Colour task, $\chi^2(9) = 19.23, p < .05$ and the Homonym Judgment task and the Synonym Judgement task, $\chi^2(9) = 34.72, p < .001$. No significant differences in performance on the Synonym Judgement task and Colour-Colour Task were identified, $\chi^2(9) = 13.03, p = .16$. However, a partial zero-order correlation revealed that performance on the Synonym judgement task and homonym judgement task were significantly correlated $r(102) = .23, p = .16$, partialling out the effects of the colour-colour task.

Bayesian contingency tables. Bayesian contingency tables (prior concentration = 1) were used to determine the difference between the alternative naming task performance, namely the Synonym Judgement task, the Homonym Judgment Task and the Colour-Colour Task. First, a Bayesian contingency table analyses revealed that there were the data were more 125 times likely under the null hypotheses (no difference in task performance) for the Synonym Judgement Task and the Colour-Colour task. Second, from the Bayes factor estimates, the evidence did not favour the alternative or null hypothesis for the difference between the Homonym judgment task

and the Colour-Colour task could not be concluded ($BF_{10} = 0.364$). Third, Bayesian contingency table analyses revealed that data were 101.375 times more likely under the alternative (a difference in task performance) for the Synonym Judgment Task and Homonym Judgment Task than the null hypothesis of no difference ($BF_{10} = 101.375$). Overall, the current Bayesian contingency tables indicate the following conclusion can be made: (1) no difference in task performance on the Synonym judgment task and the Colour-Colour task is most likely (2) a difference task performance on the Synonym judgment task and the Homonym Judgment task is most likely.

Discussion

The present study investigated the relations between false belief understanding, metalinguistic awareness, false sign understanding, and conflict inhibition. The aim of the current work was to evaluate a variety of theories—Theory of Mind Mechanism theory, Representational Processing Theory, the Executive Function theory and the Representational Content Theory—to clarify the role, if any, of metarepresentation in false belief understanding. Our objective was to examine whether false belief understanding relies on a domain-specific process unique to reasoning about mental representations, or a domain-general reasoning about both mental and non-mental representations. Further, if the evidence indicated a conceptual change in metarepresentational understanding, the present study aimed to elucidate whether the content of the representation (pictorial versus propositional) was differentially linked to the false belief task performance. The final objective was to investigate the role of conflict inhibition in false belief, false sign and metalinguistic task performance.

As a result of the lack of significant correlations between the false belief tasks and the non-mental representation tasks (metalinguistic tasks and false sign tasks) the proposed models could not be evaluated directly. The study predictions will therefore be evaluated more generally by discussing the role of domain-specific processes in false belief understanding, as well as the role of conflict inhibition in task performance. Additionally, the hypothesis regarding differences in children's performance on the alternative naming tasks, and age-related effects are reviewed.

Theory of Mind Mechanism and Domain-Specific Processes

Contrary to our predictions, the current data indicate that false belief reasoning may make unique (domain-specific) processing demands. We found children's performance on the false belief tasks were not significantly correlated with any other task (including, metalinguistic tasks,

false sign tasks, conflict inhibition tasks, or expressive vocabulary). Additionally, Bayesian regression analyses revealed that the combination of these tasks in predicting false belief task performance provided stronger evidence for null hypothesis (no effect), in comparison to the alternative hypothesis. These findings provide preliminary evidence for the Theory of Mind Mechanism Theory.

To review, proponents of the Theory of Mind Mechanism theory suggest that metarepresentational processes are unique to reasoning about mental representations, and all other related cognitive processes play an auxiliary role in false belief understanding (Cohen & German, 2010; Cohen, German & Sasaski, 2015; Frith & Frith, 2003; Leslie, 1994; Leslie, Friedman, & German, 2004; Leslie & Thaiss, 1992; Saxe et al., 2004; Scholl & Leslie, 1999; Scott & Baillargeon, 2009). The lack of significant correlations between the false belief tasks, the non-mental representation tasks (metalinguistic tasks and false sign tasks) as well as the control tasks (conflict inhibition and expressive vocabulary) suggests that there may be something unique about reasoning about mental representations. In other words, the findings align with Cohen and colleagues (2015) suggestion there may be a specialized mechanism for theory of mind reasoning, which differs from reasoning about non-mental representations: pictorial (e.g., false signs) or linguistic (e.g., metalinguistic) representations. The lack of uniform performance on these representational tasks suggests there is no coherent construct of metarepresentation. Therefore, it may be plausible that each cognitive task requires unique processing demands.

Following from Carpendale, Hammond and Atwood (2013) paper, our findings suggest that theory of mind may develop in a bottom-up fashion. Bottom-up development, as applied to theory of mind, means that children develop an understanding of other minds within a specific

context, which expands over the course of development. Therefore, false belief understanding should be understood within a broader developmental context rather than in a vacuum (Carpendale & Müller, 2012). A key facet of this developmental systems approach is the notion that constant interaction between an individual's biology and context facilitates social development. For this reason, the developmental systems framework cautions against ascribing complex cognitive processes (i.e., using a top-down method) to explain a set of related skills when conceptualizing developmental change. For example, attributing false belief understanding to a domain-general conceptual change in representational understanding would fail to acknowledge the complexity in context that precede the development of this skill. That is, inferring that false belief understanding is based on representations assumes prior knowledge, but does not explain this knowledge. By contrast, the developmental systems approach suggests that the only way to derive meaning from developmental processes (e.g., theory of mind development) is through an understanding of system of social interactions. Therefore, the lack of significant associations found between false belief task performance and the other measures suggests that perhaps a unique development context lead to the emergence of these skills.

Our preliminary evidence for domain-specific processes in false belief task performance is also consistent with findings from neuroimaging studies. Neuroimaging studies suggest that right temporal parietal juncture (rPTJ) is selectively responsive to false belief tasks performance in both child and adult populations (Sabbagh & Taylor, 2000; Sabbagh, Bowman, Evraire & Ito 2009; Saxe & Wexler, 2005). By contrast, the left temporal parietal juncture appears to be activated for both false belief and false sign tasks (Aichhorn et al., 2009). Activation in these neural regions was found to remain when controlling for executive function and age. In line with this evidence, our findings suggest that false belief reasoning may be require a unique set of

neural processes that are not explained entirely by the cognitive demands of the task itself. Specifically, we found that our data strongly favoured the null hypothesis (no effect) for the combination of representational task (metalinguistic tasks and false sign tasks) and conflict inhibition (executive function) on false belief task performance. Taken together, these findings suggest that there may be a specific neural region dedicated to processing mental representations. However, it is important to note that the activation of the rTPJ does not imply the neural circuit is specific to this task.

Evidence against the representational processing theory.

Our prediction that mental and non-mental representational tasks would develop simultaneously due to a shared underlying conceptual change in metarepresentational understanding (i.e., Representational Processing Theory) was not supported. Our predictions were made based on the findings of Doherty, Perner and colleagues (Doherty, 2000; Doherty & Perner, 1998; Perner et al., 2002; Perner & Leahy, 2018), which identified strong correlations ($.53 \leq r \leq .72$) between false belief and metalinguistic task performance (including the Synonym Judgement Task and the Homonym Judgement Task). The current work did not replicate these findings; the correlations between False belief understanding and metalinguistic task performance were found to be non-significant ($-.04 \leq r \leq .09$). There are a few important possible explanations for the lack of replication. First, the task script for the Synonym and Homonym Task were modified from the original version in the current work. The original script proceeds as follows:

“This is Tony, and he is going to tell us what these things are. But Puppet also wants to play but he doesn’t want to say the same thing as Tony, so we are going to help him think of a different way of saying what these things are. Now, Tony says this is a rabbit. What could puppet say it is? I know, he could say it is a bunny, because it is a rabbit and it is a bunny. They are different ways of saying the same thing (Doherty, 1994).”

The revised script proceeds as follows:

“This is my friend Goofy, and he’s going to play this game with us! This can be called a bunny or a rabbit – what would you like to call it? (child responds). Now in this game it’s Goofy’s job to say the other name, not the name you said. *Remember* you said (child’s response). So, what should Goofy say? [corrections provided]”

Our modified script reduces the language (e.g., shorter in length) and working memory demands (e.g., the reminders included) of the task, which may have made success on these tasks considerably easier in the current work. Additionally, these instructions require the child to actively engage with the task as opposed to passively receiving instructions. Second, the associations between the Synonym/Homonym judgement task and false belief tasks, therefore, are mostly restricted to the work conducted by the authors of the original task (see, Doherty & Perner, 1998). Garnham, Brooks, Garnham, and Ostefeld (2000) attempted to replicate these findings, but only found significant correlations between homonym understanding (but not synonym) and false belief understanding and suggested this association was due to the shared inhibitory demands of the task; however, executive function was not controlled for in this study. Hacin (2016) replicated the association between theory of mind and metalinguistic awareness with Slovenian speaking children; yet, different metalinguistic tasks were used Hacin (2016) study and the children in this study were older (4;6 to 6 years) than the children in the current work, which could explain these differences. However, importantly, the results of the Bayesian correlational analysis (and Bayesian Partial Correlations) suggest that there is insufficient evidence to draw conclusions regarding a correlation between the false belief and metalinguistic tasks. Thus, with additional data, the original findings may be replicated. It is tentatively concluded that the false belief and metalinguistic tasks do not rely on a shared conceptual change in representational understanding.

Additionally, our prediction that false belief task performance would be significantly associated with false sign task performance was not supported. Although a small significant correlation ($r = .2$) was identified for the association between false belief contents task and the false sign location task, Bayesian analyses indicated the data was not sensitive enough to make conclusions regarding an association. Further, partialling out the effects of conflict inhibition and expressive vocabulary, this association was no longer significant. This finding aligns with that of Sabbagh and colleagues (2006; Experiment 2) that the association between false belief and false sign task performance became non-significant when age was partialled out. Therefore, these findings suggest that previously identified associations between these tasks may be better attributed to shared conflict inhibition demands, instead of the ability to metarepresent.

Interestingly, the non-mental representational tasks (false sign and metalinguistic task) were also not significantly correlated. To our knowledge, no previous studies have included both false sign and metalinguistic tasks in the same study. The percentage of correct response on the false sign tasks (contents: 62.5%; location: 72.12%) was generally higher than the percent of correct responses on the metalinguistic tasks (synonym: 55.70%; homonym: 35.70%). Perhaps, the pictorial content of the representation explains the discrepancy between these non-mental representational tasks (see, Cohen, German & Sasaski, 2015). In general, the lack of correlation between these non-mental representational tasks supports the Theory of Mind Mechanism theory that there is a lack of shared metarepresentational processes in these tasks. Again, it is important to note, Bayesian correlation analysis suggest there is insufficient data to draw conclusions regarding the association between these tasks.

Overall, the current findings suggest that metarepresentation is not a coherent construct that contributes to reasoning about both mental (false beliefs) and non-mental representations

(metalinguistic tasks and false sign tasks). Our findings provide preliminary evidence for the Theory of Mind Mechanism theory: a unique domain-specific process for false belief reasoning.

The Role of Conflict inhibition

Contrary to our predictions, conflict inhibition was not correlated with false belief task performance. These predictions were made based on past research, which demonstrates moderate associations between conflict inhibition and false belief understanding (see meta-analysis, Devine & Hughes, 2014). This lack of correlation could be attributed to the general findings which suggest that executive function (e.g., conflict inhibition) does not account for the complete variance in false belief understanding. For example, Sabbagh, Xu, Carlson, Moses and Lee (2006) found that Chinese children had significantly better executive function than American children, yet both children performed similarly on false belief tasks. Therefore, the lack of correlation may be due to the unique domain-specific cognitive processes involved in theory of mind over and above executive function (or, conflict inhibition). It is also possible that our evidence was not strong enough to detect an effect (for Bayes Factor interpretation see, Dienes, 2014).

Additionally, our hypothesis that conflict inhibition would predict false belief task performance was not supported. The results of our Bayesian regression analysis indicate that conflict inhibition (with combined effects of age [in months] and expressive vocabulary) was more likely under the null hypothesis (no effect) in predicting false belief task performance. Our predictions were made based on previous research which suggests that executive function significantly predicts false belief task performance in young children both concurrently (Müller, Jacques, Brocki, & Zelazo, 2009) and prospectively (Hughes & Ensor, 2007; Marcovitch et al., 2015; Müller, Liebermann-Finestone, Carpendale, Hammond, & Bibok, 2012). There are a few

potential explanations for the absence of an effect in the current work. First, our age range may have been too narrow to detect a developmental progression on these tasks. Perhaps including an age range from two to five years would better capture this developmental change. Second, our sample was pulled from a high socioeconomic status area which may have affected children's performance on both the false belief (see, meta-analysis Devine & Hughes, 2018) and conflict inhibition (Lawson, Hook & Farah, 2018) measures. Finally, Bayesian regression analyses indicated that there was insufficient evidence to draw conclusions regarding an effect of conflict inhibition (independently) on false belief task performance. Therefore, age and expressive vocabulary may account for a significant amount of variation false belief task performance contributing to the null findings. Given the current findings, no conclusions can be drawn regarding the role of conflict inhibition in false belief understanding.

Conflict inhibition was significantly correlated with both non-mental representational tasks: synonym/homonym tasks and false sign tasks. To our knowledge, previous research has not measured conflict inhibition directly when evaluating synonym and homonym understanding. However, our findings are consistent with Garnham and colleagues' (2000) suggestion that the association between false belief and metalinguistic task performance is better explained by the executive function demands of the task than representational processing. The significant associations found between conflict inhibition and false sign task performance are consistent with the findings by Sabbagh, Moses and Shiverick (2006; study 2). As Sabbagh and colleagues suggest, conflict inhibition may be important to understanding false signs because it enables children to consider is a true sign representation (prototypical sign-use) may be false. These findings suggest that conflict inhibition may be a central cognitive skill required to reason about synonyms/homonyms and false signs.

The current work suggests that conflict inhibition is not correlated with false belief task performance, but is correlated with the synonym/homonym tasks and the false sign task. Taken together these findings suggest that theory of mind reasoning may require unique processing demands, over and above conflict inhibition.

Alternative Naming Task Performance

Our hypothesis that the Colour-Colour task would be significantly easier than the Synonym and Homonym Judgement task was not supported. The Colour-Colour task has been included as a control in previous work because it is presumed to have similar conflict inhibition demands as the metalinguistic tasks, but does not require the ability to represent objects in multiple ways. Interestingly, we did not find a correlation between the Colour-Colour task and the conflict inhibition tasks; however, both the Synonym/Homonym judgement task were significantly correlated with the conflict inhibition tasks. Thus, Perner, Stummer, Sprung and Doherty (2002) suggestion that the Colour-Colour task has the same conflict inhibition demands as the Synonym/Homonym Judgement task was not supported. Additionally, contrary to Perner, and colleagues (2002) findings, we did not find a significant difference between the Colour-Colour task and the Synonym Judgement Task. Therefore, it may be inferred that the executive function demands of these tasks, rather than representational demands explains their association.

Additionally, the Homonym Judgment task was significantly harder for children than both the Colour-Colour and the Synonym Judgment Task, respectively. These findings are consistent with Doherty's (2004) argument that children show specific difficulties understanding the meaning of homonyms (secondary meanings), which may persist until age ten. Children's specific difficulty with the Homonym Judgment task suggests this task may require unique

processing demands beyond conflict inhibition (e.g., Colour-Colour) and the ability to metarepresent (e.g., Synonym Judgment Task).

Age-Related Effects

The prediction of age-related effect on false belief, metalinguistic and false sign task performance was not supported. Our findings suggest the alternative hypothesis, or an effect of age is more likely than the null hypothesis for the children's performance false sign task performance, but not for their performance on the false belief or metalinguistic tasks. The false sign tasks were considerably easier for children to pass (higher overall percent of pass rate) than the false belief and metalinguistic tasks. As DeLoache (2004) suggests children are able to understand symbol-referent relations—which is required for the false sign task performance—earlier (around two-and-a-half years) than more abstract (e.g., mental) symbols. Children typically pass false belief (see, Wellman et al., 2001) and metalinguistic tasks (Doherty & Perner, 1998; Doherty, 2000; Garnham et al., 2002; Perner & Leahy, 2015) around four years of age. Therefore, the three-to-four age range of the current study may have only captured the understanding of signs (pictorial symbols), and not linguistic or mental representations. Overall, the narrow age range of the current work could explain the lack of age effects on these core tasks.

Limitations of the Current Study

The current study has several limitations that may have contributed to the findings. The most significant limitations were the size and characteristics of our sample. A priori power analysis revealed that our sample size was sufficient ($n = 104$) for the frequentist analyses; however, the Bayesian Factor analysis suggests that many of the correlational and regression analysis were not sensitive enough to detect an effect implying a need for further data collection.

With a larger sample we would have been able to draw more definitive conclusions regarding the alternative or null hypotheses associated with our proposed models. The narrow age range (3;2 to 4;7 years) was selected for the current work because children typically pass the false belief tasks (Wellman, Cross & Watson, 2001), metalinguistic tasks (Edwards & Kirkpatrick, 1999) and false sign tasks around four years. Yet, this age range may have been too restricted to identify developmental progression on these tasks. Additionally, the Victoria community sample does not represent a diverse sample population. Our participants were drawn from high socioeconomic status (SES) areas. Research shows that SES does play a role in children's theory of mind development (see, meta-analysis Devine & Hughes, 2018), suggesting a potential bias in our sample. Given the limitations of our sample results should be interpreted cautiously.

In addition to the limitations of our sample, the measures used in the current work also had limitations. First, the current work had a variety of ceiling and floor effects on the measures may have contributed to the lack of variability in the results. The floor effects on false belief measure (the dependent variable), and ceiling effects on some of the independent variables (the Synonym Judgment task and false sign tasks) may have skewed our results. Second, the current work only used conflict inhibition as a measure of executive function; however, other aspects of executive function, namely working memory and cognitive flexibility, may have played a role in children's task performance.

A final limitation of this study is the cross-sectional design. As a result of this design, we are unable to examine directionality in the relationship between false belief understanding, metalinguistic awareness, false sign understanding and conflict inhibition.

Future Directions

Although the results show preliminary evidence the metarepresentation may be unique to false belief reasoning, further research is required. Specifically, future research should address the limitations of the current work, and extend the current findings using neuroimaging studies, conflict inhibition controls and a longitudinal design.

The limitations of our sample should be addressed in future work. The current findings should be replicated with a larger sample to allow for additional conclusions to be drawn from the Bayesian analyses drawn. Additionally, diversifying the sample and using a less restrictive age-range will make the findings more definitive. Expanding the age range may also improve the variability in our results, and reduce both floor and ceiling effects on our measures. Introducing a variety of measure for each construct (e.g., more than two measures of false belief understanding) would also likely improve the variability of the results.

Future research should consider a confirmatory factor analysis (CFA) to determine if related tasks are measuring the same underlying latent construct. For example, using a CFA to determine if the false belief location, and false belief contents tasks are both measuring the same latent construct: theory of mind. Using a CFA for the false sign tasks and metalinguistic tasks would also be beneficial.

Neuroimaging studies are recommended to provide more additional sources of evidence for the arguments made by proponents of the Theory of Mind Mechanism Theory. To our knowledge, no study has looked at the neural correlates of false belief in comparison to processing both propositional (Synonym/Homonym understanding) or pictorial (false sign understanding) representational understanding. This neuroimaging work could clarify the role of the rPTJ in representational processing.

Further clarification of the role of conflict inhibition in reasoning about both mental and non-mental representations should be considered. Our findings provide preliminary evidence for the importance of conflict inhibition in non-mental representation tasks. However, future research should test the associations between representational tasks (e.g., linguistic, pictorial, photographs) controlling for conflict inhibition. Controlling for conflict inhibition will allow for the conclusions to be made about the representational processing theory: does executive function explain all the variance in the associations between (mental and non-mental) representational tasks?

Finally, future research should conduct a longitudinal study to establish the directionality of the relationship between false belief tasks, metalinguistic task, false sign tasks and executive function. Past research has yet to explore the prospective relations between these variables, which may provide interesting insight into the mechanisms that contribute to the ability to reason about mental and non-mental representations.

Conclusion

The aim of the present study was to investigate the role of metarepresentational understanding and conflict inhibition in theory of mind development. In particular, we were interested in determining whether domain-specific processes or a domain-general conceptual change in representational understanding contributed to false belief understanding. We found, contrary to previous research (see, Perner et al., 2002; Sabbagh, Moses & Shiverick, 2006), that non-mental representational tasks (metalinguistic tasks and false sign tasks) were not correlated with, nor predicted false belief task performance. These results suggest that domain-specific mechanisms may contribute to false belief understanding, providing preliminary evidence for the Theory of Mind Mechanism Theory. Specifically, these findings provide evidence for a bottom-

up model of the development of theory of mind, and unique processing demands that contribute to the development of each cognitive skills. Conflict inhibition was found to be significantly associated with the non-mental representation tasks highlighting the importance of the executive function for the development of these cognitive skills. Future longitudinal research investigating the directionality in the relations between false belief understanding and non-mental representational tasks should be conducted over the preschool period to clarify the role of metarepresentation in theory of mind development.

References

- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporoparietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. *Journal of Cognitive Neuroscience*, 21, 1179–1192. doi: 10.1162/jocn.2009.2108
- Apperly, I. A., Samson, D., Chiavarino, C., Bickerton, W. L., & Humphreys, G. W. (2007). Testing the domain-specificity of a theory of mind deficit in brain-injured patients: Evidence for consistent performance on non-verbal, "reality-unknown" false belief and false photograph tasks. *Cognition*, 103(2), 300-321. doi:10.1016/j.cognition.2006.04.012
- Astington, J. W. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35, 1311–1320. doi:10.1037/0012-1649.35.5.1311
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118. doi:10.1016/j.tics.2009.12.006
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, 21, 37–46. doi: 10.1016/0010-0277(85)90022-8
- Bernstein, D. M., Coolin, A., Fischer, A. L., Thornton, W. L., & Sommerville, J. A. (2017). False-belief reasoning from 3 to 92 years of age. *PLoS One*, 12(9), e0185345. doi:10.1371/journal.pone.0185345
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77(1), B25-B31. doi:10.1016/S0010-0277(00)00096-2
- Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the

- mind or predicting signal-dependent action? performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development*, 6(2), 259-283. doi:10.1207/s15327647jcd0602_5
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol 53, pp. 301–344). Academic Press: Burlington
- Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. *Child Development*, 67, 836–849. doi:10.1111/j.1467-8624.1996.tb01767.x
- Brownell, C. A., Svetlova, M., Anderson, R., Nichols, S. R., & Drummond, J. (2013). Socialization of early prosocial behavior: Parents' talk about emotions is associated with sharing and helping in toddlers. *Infancy*, 18, 91–119. doi:10.1111/j.1532-7078.2012.00125.x6;
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337-342. doi:10.1016/j.cognition.2009.05.006
- Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology*, 119, 120. doi:10.1016/j.jecp.2013.10.002
- Carlson, E., Jenkins, F., Li, T., & Brownell, M. (2013). The interactions of vocabulary, phonemic awareness, decoding, and reading comprehension. *The Journal Of Educational Research*, 106(2), 120-131. doi:10.1080/00220671.2012.687791
- Carlson, S. M., Claxton, L. J., & Moses, L. J. (2013). The relation between executive function

- and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*, 186–197. doi:10.1080/15248372.2013.824883
- Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(4), 391–402. doi:10.1002/wcs.1232
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053. doi:10.1111/1467-8624.00333
- Carpendale, J. I. M., Hammond, S. I., & Atwood, S. (2013). A relational developmental systems approach to moral development. (pp. 125-153). United States: Elsevier Science & Technology. doi:10.1016/B978-0-12-397946-9.00006-36
- Carpendale, J. I. M., & Müller, U. (2012). Flaws in the notion of rational imitation: On the impossible leap from computation to rationality and normativity. *Human Development, 55*(3), 122-128. doi:10.1159/000339720
- Carrington, S., & Bailey, A. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30*(8), 2313-2335. doi:10.1002/hbm.20671
- Chaney, C. (1994). Language development, metalinguistic awareness, and emergent literacy skills of 3-year-old children in relation to social class. *Applied Psycholinguistics, 15*(3), 371–394. doi:10.1017/S0142716400004501
- Cervantes, C. A., & Callanan, M. A. (1998). Labels and explanations in mother-child emotion talk: Age and gender differentiation. *Developmental Psychology, 34*, 88–98. doi:10.1037/0012-1649.34.1.88
- Christ, S. E., Holt, D. D., White, D. A., & Green, L. (2007). Inhibitory control in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*(6), 1155-

1165. doi:10.1007/s10803-006-0259-y

- Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition, 115*(3), 417-425. doi:10.1016/j.cognition.2010.03.001
- Cohen, A. S., Sasaki, J. Y., & German, T. C. (2016). Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations? *Cognition, 136*, 49-63. doi:10.1016/j.cognition.2014.11.016
- de Villiers, J. G., & de Villiers, P. A. (2003). Language for thought: Coming to understand false beliefs. *Language in mind: Advances in the study of language and thought*, 335-384.
Retrieved from: <http://isites.harvard.edu/fs/docs/icb.topic151491.files/devillers-paper.pdf>
- Deleau, M. (2012). Language and theory of mind: Why pragmatics matter. *European Journal of Developmental Psychology, 9*(3), 295-312. doi:10.1080/17405629.2012.680303
- DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science, 238*(4833), 1556-1557. doi:10.1126/science.2446392
- DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences, 8*(2), 66-70. doi:10.1016/j.tics.2003.12.004
- DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition, 52*(2), 83-110. doi:10.1016/0010-0277(94)90063-9
- DeLoache, J. S., Miller, K. F., & Rosengren, K. S. (1997). The credible shrinking room: Very young children's performance with symbolic and nonsymbolic relations. *Psychological Science, 8*(4), 308-313. doi:10.1111/j.1467-9280.1997.tb00443.x
- DeLoache, J. S., & Smith, C. M. (1999). Early symbolic representation. In I. E. Sigel

(Ed.), *Development of mental representation: Theories and applications* (pp. 61-86).

Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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

doi:10.1111/cdev.12237

Devine, R. T., & Hughes, C. (2018). Family correlates of false belief understanding in early childhood: A meta-Analysis. *Child Development, 89*(3), 971-987.

doi:10.1111/cdev.12682

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language, 27*(2), 367–392.

doi:10.1017/S0305000900004153

Doherty, M. J. (2002). Synonyms, homonyms and false beliefs: A reply to Garnham, Brooks, Garnham and Ostefeld. *Developmental Science, 5*, 84–85. doi:10.1111/1467-7687.00212

Doherty, M. (1994). Metalinguistic understanding and theory of mind. Unpublished doctoral dissertation, Laboratory of Experimental Psychology, University of Sussex.

Doherty, M. J., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development, 13*, 279–305. doi:10.1016/S0885-

2014(98)90012-0

Dunn, J., & Cutting, A. L. (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development, 8*, 201–219. doi:10.1111/1467-

9507.00091

Dunn, J., Cutting, A. L., & Demetriou, H. (2000). Moral sensibility, understanding others, and

- children's friendship interactions in the preschool period. *British Journal of Developmental Psychology*, 18(2), 159-177. doi:10.1348/026151000165625
- Edwards, H. T., & Kirkpatrick, A. G. (1999). Metalinguistic awareness in children: A developmental progression. *Journal of Psycholinguistic Research*, 28, 313–329. doi:10.1023/A:1023275214000
- Egeth, M., & Kurzban, R. (2009). Representing metarepresentations: Is there theory of mind-specific cognition? *Consciousness and Cognition*, 18(1), 244-254. doi:10.1016/j.concog.2008.07.005
- Etz, A., & Vandekerckhove, J. (2016). A bayesian perspective on the reproducibility project: Psychology. *PloS One*, 11(2), e0149794. doi:10.1371/journal.pone.0149794
- Farrar, M. J., & Ashwell, S. (2012). Phonological awareness, executive functioning, and theory of mind. *Cognitive Development*, 27, 77–89. doi:10.1016/j.cogdev.2011.08.002
- Farrar, M. J., Ashwell, S., & Maag, L. (2005). The emergence of phonological awareness: Connections to language and theory of mind development. *First Language*, 25(2), 157–172. doi:10.1177/0142723705051980
- Fernández, C. (2013). Mindful storytellers: Emerging pragmatics and theory of mind development. *First Language*, 33, 20–46. doi:10.1177/0142723711422633
- Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing Theories of Mind*, New York, NY: Press Syndicate of the University of Cambridge (pp. 244–267).
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1989). A transitional period in the development of

- the appearance-reality distinction. *International Journal of Behavioral Development*, 12, 509–526. doi:10.1177/016502548901200407
- Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive Development* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Fletcher, P. C., Happé, F., Frith, U., Frith, C. D., Baker, S. C., Dolan, R. J., & Frackowiak, R. S. J. (1995). Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition*, 57(2), 109-128. doi:10.1016/0010-0277(95)00692-R
- Flood, J., & Salus, M. W. (1982). Metalinguistic awareness: Its role in language development and its assessment. *Topics in Language Disorders*, 2(4), 56-64. doi:10.1097/00011363-198209000-00009
- Frege, G. (1948). Sense and reference. *The philosophical review*, 57(3), 209-230. Retrieved from: <http://www.jstor.org/stable/2181485>
- Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Psychological Science*, 10(5), 151-155. Retrieved from: <http://journals.sagepub.com/doi/pdf/10.1111/1467-8721.00137>
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 459-473. doi:10.1098/rstb.2002.1218
- Froese, T. (2012). From adaptive behavior to human cognition: A review of enaction. *Adaptive Behavior*, 20(3), 209-221. doi:10.1177/1059712311433892
- Garnham, A., & Garnham, W. A. (2002). Metarepresentation or inhibition? An open question: A response to Doherty. *Developmental Science*, 5, 86–86. doi:10.1111/1467-7687.00213
- Garnham, W. A., Brooks, J., Garnham, A., & Ostefeld, A. (2000). From synonyms to

- homonyms: Exploring the role of metarepresentation in language understanding.
Developmental Science, 3, 428–441. doi:10.1111/1467-7687.00137
- Goldman, A. I. (2012). Theory of mind. *The Oxford handbook of philosophy of cognitive science*, 402-424.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59, 26–37. doi:10.1111/1467-8624.ep8860955
- Gopnik, A., Kuhl, P. K. 1., & Meltzoff, A. N. (1999). *The scientist in the crib: Minds, brains, and how children learn*. New York: William Morrow & Co.
- Gopnik, A., & Wellman H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(2), 145–171. doi:10.1111/j.1468-0017.1992.tb00202.x
- Guan, Y., & Farrar, M. J. (2016). Do chinese- and english-speaking preschoolers think differently about language? *First Language*, 36(6), 617-636.
doi:10.1177/0142723716673956
- Hacin, K. (2016). Relationship between children's theory of mind and metalinguistic competence in early childhood. *Psihološka Obzorja*, 25(72–83). doi:10.20419/2016.25.448
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6(3), 346–359. doi:10.1111/1467-7687.00289
- Happaney, K. R., & Zelazo, P., (2008). Less specificity in higher cognitive mechanisms: Evidence from theory of mind. In M. J., Roberts, (Ed.). *Integrating the mind: Domain general versus domain specific processes in higher cognition*. (pp. 257 - 273). Psychology Press.
- Harris, P. L. (2006) Social cognition. In: D. Kuhn, R. Siegler (Eds.), *Handbook of Child*

- Psychology* (811–858). New York, NY: John Wiley & Sons. doi:10.1002/9780470147658
- Harris, P. L., Rosnay, M. D., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, *14*(2), 69–73. doi:10.1111/j.0963-7214.2005.00337.x
- Hartwright, C., Apperly, I., & Hansen, P. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, *67*, 183-192. doi:10.1016/j.neuropsychologia.2014.12.015
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360-1366. doi:10.1126/science.1146282
- Hughes, C., & Dunn, J. (1997). “Pretend you didn't know”: Preschoolers' talk about mental states in pretend play. *Cognitive Development*, *12*(4), 477–497. doi:10.1016/S0885-2014(97)90019-8
- Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447-1459. doi:10.1037/0012-1649.43.6.1447
- Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of Mind and Prosocial Behavior in Childhood: A Meta-Analytic Review. *Developmental Psychology*, *52*(8), 1192-1205. doi: 10.1037/dev0000140
- Jastrow, J. (1899). The mind's eye. *Popular Science Monthly*, *54*, 299-312.
- Jeffreys, H. (1961). *Theory of probability*, (3rd ed.) Oxford, UK: Oxford University Press.
- Juan, V. S., & Astington, J. W. (2012). Bridging the gap between implicit and explicit

- understanding: How language development promotes the processing and representation of false belief. *British Journal of Developmental Psychology*, 30(1), 105-122. doi: 10.1111/j.2044-835X.2011.02051.x
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test*. John Wiley & Sons, Inc.
- Kim, S., (2015). ppcor: Partial and Semi-Partial (Part) Correlation. R package version 1.1. <https://CRAN.R-project.org/package=ppcor>
- Kirk, E., Pine, K., Wheatley, L., Howlett, N., Schulz, J., & Fletcher, B. (2015). A longitudinal investigation of the relationship between maternal mind-mindedness and theory of mind. *British Journal of Developmental Psychology*, 33(4), 434–445. doi:10.1111/bjdp.12104
- Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, 21(2), e12529-n/a. doi:10.1111/desc.12529
- Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, 40(3), 203-218. doi:10.1016/0010-0277(91)90025-Y
- Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal Of Developmental Psychology*, 26(4), 485-497. doi:10.1348/026151007X260154
- Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological review*, 94(4), 412. doi: 10.1037/0033-295X.94.4.412
- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50(1), 211-238. doi:10.1016/0010-0277(94)90029-9

- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences*, 8(12), 528-533. doi:10.1016/j.tics.2004.10.001
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43(3), 225-251. doi:10.1016/0010-0277(92)90013-8
- Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. *Cognitive, Affective & Behavioral Neuroscience*, doi:10.3758/s13415-018-0568-6
- Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44(2), 523-531. doi:10.1037/0012-1649.44.2.523
- Lohmann, H., & Tomasello, T. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, 74, 1130–1144. doi:10.1111/1467-8624.00597
- Losada, A. (2016). Mental files, what for? *Review of Philosophy and Psychology*, 7(2), 405-419. doi:10.1007/s13164-015-0284-x
- 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. doi:10.1016/j.cogdev.2014.07.001
- McAlister, A., & Peterson, C. (2007). A longitudinal study of child siblings and theory of mind development. *Cognitive Development*, 22(2), 258–270. doi:10.1016/j.cogdev.2006.10.009
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A.,

- & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes knees-shoulders task. *Frontiers in Psychology*, *5*, 1–14. doi:10.3389/fpsyg.2014.00599
- Miller, S. A., (2012). *Theory of mind: Beyond the preschool years*. New York: Psychology Press.
- 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. doi:10.1111/j.1467-8624.2007.01018.x
- Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*(3), 722–730. doi:10.1111/j.1467-8624.1990.tb02815.x
- Morey, R., & Rouder, J., (2015). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-2. <https://CRAN.R-project.org/package=BayesFactor>
- Moses L. J., & Sabbagh, M. A., (2008). Interactions between domain general and domain specific processes in the development of children's theory of mind. In M. J., Roberts, (Ed.). *Integrating the mind: Domain general versus domain specific processes in higher cognition*. (pp. 275 - 291). Psychology Press.
- Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I. M., 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. doi: 10.1016/j.jecp.2011.08.014
- Müller, U., Miller, M. R., Michalczyk, K., & Karapinka, A. (2007). False belief understanding:

- The influence of person, grammatical mood, counterfactual reasoning and working memory. *British Journal of Developmental Psychology*, 25(4), 615-632.
doi:10.1348/026151007X182962
- Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development*, 20(2), 173-189.
doi:10.1016/j.cogdev.2004.12.004
- Nelson, K., Plesa Skwerer, D., Goldman, S., Henseler, S., Presler, N., & Fried Walkenfeld, F. (2003). Entering a community of minds: an experiential approach to 'theory of mind'. *Human Development*, 46, 24-46. doi: 10.1159/000067779
- Nuijten, M., Wetzels, R., Matzke, D., Dolan, C. & Wagenmakers, E. J., (2015). BayesMed: Default Bayesian Hypothesis Tests for Correlation, Partial Correlation, and Mediation. R package version 1.0.1. <https://CRAN.R-project.org/package=B>
- Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology*, 20(2), 205-224.
doi:10.1348/026151002166415
- Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, 48(12), 1243-1250. doi:10.1111/j.1469-7610.2007.01810.x
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or asperger syndrome. *Child Development*, 83(2), 469-485.
doi:10.1111/j.1467-8624.2011.01728.x
- Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

- Perner, J. (2000). RUM, PUM, and the perspectival relativity of sortals. *Minds in the making: Essays in honor of David R. Olson* (pp. 212–232). Oxford, UK: Oxford Press.
- Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60*(3), 689–700. doi:10.1111/j.1467-8624.1989.tb02
- Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intentionality. *Cognition, 145*, 77–88. doi:10.1016/j.cognition.2015.08.006
- Perner, J., & Leahy, B. (2016). Mental files in development: Dual naming, false belief, identity and intensionality. *Review Of Philosophy And Psychology, 7*(2), 491–508. doi:10.1007/s13164-015-0235-6
- Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science, 333*(6041), 474–477. doi:10.1126/science.1201216
- Perner, J., Stummer, S., & Lang, B. (1999). Executive functions and theory of mind: Cognitive complexity or functional dependence? In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 133–152). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development, 17*(3), 1451–1472. doi: 10.1016/S0885-2014(02)00127-2
- Piekny, J., Grube, D., & Maehler, C. (2013). The relation between preschool children's false-belief understanding and domain-general experimentation skills. *Metacognition and Learning, 8*(2), 103. doi:10.1007/s11409-013-9097-4

- Pillow, B. H. (1999). Epistemological development in adolescence and adulthood: A multidimensional framework. *Genetic, Social, and General Psychology Monographs*, *125*(4), 413. Retrieved from:
<http://web.a.ebscohost.com.ezproxy.library.uvic.ca/ehost/detail/detail?vid=6&sid=30307c1e-0d65-4232-a509792f0ba8732%40sessionmgr4>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526. doi:10.1017/S0140525X0007651
- Proust, J. (2007). Metacognition and metarepresentation: Is a self-directed theory of mind a precondition for metacognition? *Synthese*, *159*(2), 271-295. doi:10.1007/s11229-007-9208-3
- Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Ve
- Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in grades 4, 6, and 8. *Journal of Experimental Child Psychology*, *102*(1), 96-113. doi:10.1016/j.jecp.2008.01.004
- Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (2009). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, *80*(4), 1147-1162. doi:10.1111/j.1467-8624.2009.01322.x
- Sabbagh, M. A., & Callanan, M. A. (1998). Metarepresentation in action: 3-, 4-, and 5-year-olds' developing theories of mind in parent-child conversations. *Developmental Psychology*, *34*(3), 491-502. doi:10.1037/0012-1649.34.3.491
- Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers'

- understanding of false beliefs, false photographs, and false signs. *Child Development*, 77(4), 1034-1049. doi:10.1111/j.1467-8624.2006.00917.x
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind a comparison of Chinese and US preschoolers. *Psychological Science*, 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Saracho, O. N. (2014a). Theory of mind: Children's understanding of mental states. *Early Child Development and Care*, 184(6), 949-961. doi:10.1080/03004430.2013.821985
- Saracho, O. N. (2014b). Theory of Mind: Understanding Young Children's Pretence and Mental States. *Early Child Development And Care*, 184(8), 1281-1294. doi: 10.1080/03004430.2013.865617
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. the role of the temporo-parietal junction in "theory of mind". *Neuroimage*, 19(4), 1835. doi:10.1016/S1053-8119(03)00230-1
- Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, 43(10), 1391-1399. doi:10.1016/j.neuropsychologia.2005.02.013
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and ‘theory of mind’. *Mind & Language*, 14(1), 131-153. doi: 10.1111/1468-0017.00106
- Schönbrodt, F. D., & Wagenmakers, E. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128-142. doi:10.3758/s13423-017-1230-y
- Schwebel, D. C., Rosen, C. S., & Singer, J. L. (1999). Preschoolers' pretend play and theory of

- mind: The role of jointly constructed pretence. *British Journal of Developmental Psychology*, 17(3), 333–348. doi:10.1348/026151099165320
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? attributing false beliefs about object identity at 18 months. *Child Development*, 80(4), 1172-1196. doi:10.1111/j.1467-8624.2009.01324.x
- Sharpe, D., & Zelazo, P. D. (2002). The foundations and development of metalinguistic knowledge. *Journal of Child Language*, 29(2), 449-488. doi:10.1017/S0305000902295343
- Smith, C. L., & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology*, 34(3), 449-468. doi:10.1016/0022-0965(82)90071-6
- Solesa-Grijak, D. (2011). Cognition and metalinguistic awareness as sine quibus non for solving verbal problems. *Studia Psychologica*, 53(1), 41. Retrieved from: http://lg5jh7pa3n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF8&rft_id=info%3Asid%2
- Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13(6), 907. doi:10.1111/j.1467-7687.2009.00946.x
- Southgate, V., & Verneti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, 130(1), doi:10.1016/j.cognition.2013.08.008
- Spataro, P., Rossi-Arnaud, C., & Longobardi, E. (2018). Are belief-based justifications associated with metalinguistic awareness? A cross-sectional study in school-age children. *Infant and Child Development*, 27(1), doi:10.1002/icd.2048

- Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia: Meta-analysis. *The British Journal of Psychiatry*, 191(1), 5-13.
doi:10.1192/bjp.bp.107.035899
- Stanley, D. (2018). *apaTables: Create American Psychological Association (APA) Style Tables*. R package version 2.0.3. <https://CRAN.R-project.org/package=apaTables>
- Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind?. *Social Neuroscience*, 1(3-4), 309-319. doi:10.1080/17470910601029221
- Suddendorf, T. (1999). Children's understanding of the relation between delayed video representation and current reality: A test for self-awareness? *Journal of Experimental Child Psychology*, 72(3), 157-176. doi:10.1006/jecp.1998.2485
- Suddendorf, T., & Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin*, 127(5), 629-650. doi:10.1037/0033-2909.127.5.629
- Swiatczak, B. (2011). Conscious representations: An intractable problem for the computational theory of mind. *Minds and Machines*, 21(1), 19-32. doi:10.1007/s11023-010-9214-y
- Tunmer, W. E., & Herriman, M. L. (1984). *The development of metalinguistic awareness: A conceptual overview*, Berlin, GE: Springer-Verlag.
- Wagenmakers, E., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Morey, R. D. (2018b). Bayesian inference for psychology. part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58-76. doi:10.3758/s13423-017-1323-7
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Morey, R. D. (2018a).

- Bayesian inference for psychology. part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35-57. doi:10.3758/s13423-017-1343-3
- Wagenmakers, E., Verhagen, J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods*, 48(2), 413-426. doi:10.3758/s13428-015-0593-0
- Walker, R. F., & Murachver, T. (2012). Representation and theory of mind development. *Developmental Psychology*, 48(2), 509-520. doi:10.1037/a0025663
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Eds.), *Blackwell handbook of childhood cognitive development* (p. 167–187). Malden, MA: Blackwell.
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.
- Wellman, H. M., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684. doi:10.1111/1467-8624.00304
- Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, 82(3), 780-792. doi:10.1111/j.1467-8624.2011.01583.x
- Wellman, H.M. & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523-541. doi:10.1111/j.1467-8624.2004.00691.x
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function

- of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128. doi:10.1016/0010-0277(83)90004-5
- Wimmer, M. C., Robinson, E. J., Koenig, L., & Corder, E. (2014). Getting the picture: Iconicity does not affect representation-referent confusion: E107910. *PLoS One*, 9(9) doi:10.1371/journal.pone.0107910
- Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35(1), 41-68. doi:10.1016/0010-0277(90)90036-J
- Zelazo, P. D., & Mueller, U. (2010). Executive function in typical and atypical development. *The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition*, 574-603. doi: 10.1111/b.9780631218418.2004.00024.x
- Zelazo, P. D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representations. *Developmental Psychology*, 35(4), 1059-1071. doi:10.1037//0012-1649.35.4.1059

Appendix A

Smarties Task (Gopnik, 1988)

Here's a Smarties box. What do you think is inside the Smarties box?

Response: _____

Let's look and see!

Open the box.

Wow! There are really pencils inside. *Show the child the pencils. Close the box.*

What is inside the box?

Circle: PENCILS SMARTIES OTHER: _____

Here is Penguin Pete and Penguin Pete has never seen inside this box before. What will penguin Pete think is in the box, smarties or pencils?

Response: PENCILS SMARTIES OTHER: _____

Score: 0 1

Appendix B

False Belief task (Wellman & Liu, 2004)

Children see a toy figure of a boy and a sheet of paper with a backpack and a closet drawn on it.

“Here’s Scott. Scott wants to find his mittens. His mittens might be in his backpack (*place image of backpack down*) or they might be in the closet (*place image of closet down*). Really, Scott’s mittens are in his backpack. But Scott thinks his mittens are in the closet.”

“So, where will Scott look for his mittens? In his backpack or in the closet?”

Child’s target answer: backpack closet

“Where are Scott’s mittens really? In his backpack or in the closet?”

Child’s control answer: backpack closet

Score: 0 1

(0 if any or all of the two questions wrong, 1 if both questions correct)

Appendix C

Synonym Judgment Task (Doherty, 2000)

Now, we are going to play another pointing game! Here we go!

Phase 1: Vocabulary Check

Sheet A: Rabbit, cup, sun, house

Can you point to the bunny?

Can you point to the rabbit? *If incorrect – repeat question.*

Cool! This has two names (point to object)– A bunny and a rabbit. _____/1

Can you point to the cup?

Can you point to the mug? *If incorrect – repeat question.*

Wow! This has two names too (point to object)– A cup and a mug. _____/1

Sheet B: T.V., Present, chair, hat

Can you point to the T.V.?

Can you point to the television? *If incorrect – repeat question.*

Wow! This has two names (point to object)– A T.V. and a television. _____/1

Can you point to the chair?

Can you point to the present?

Can you point to the gift? *If incorrect – repeat question.*

Wow! This has two names (point to object)– A present and a gift. _____/1

Sheet C: Lady, street, spoon, monkey

Can you point to the spoon?

Can you point to the street?

Can you point to the road? *If incorrect – repeat question.*

Look at that! This has two names (point to object)– A street and road. _____/1

Can you point to the woman?

Can you point to the lady? *If incorrect – repeat question.*

Awesome! This has two names (point to object)– A woman and lady. _____/1

Score: ____/6

Phase 2: Modelling

(Put on puppet) **This is my friend, Goofy, and Goofy is going to play the game with us. Here we go!**

This can be called a Rabbit or a bunny. What would you like to call it?

Response: _____

Now, in this game, it is Goofy's job to say the other name NOT the one that you said. Let's practice. Remember you said *repeat child's response*. So what should Goofy say?

Goofy should say (OTHER NAME), because (OTHER NAME) is another name for this (*point to bunny*) but not the name you called it.

Goofy should not say (SAME NAME) because that's what you called it.

And, Goofy should not say HOUSE, because this (*point to bunny*) is not a house.

Phase 3: Testing

Okay – Let's try some more! Remember it is Goofy's job to say the other name NOT the name you said.

What would you call this? *Circle:* PRESENT GIFT

Goofy says (OTHER NAME) – Is that what Goofy should've said?

Y _____ N _____

What would you call this? *Circle:* CUP MUG

Goofy says CHAIR– Is that what Goofy should've said?

Y _____ N _____

What would you call this? *Circle:* WOMAN LADY

Goofy says (OTHER NAME) – Is that what Goofy should've said?

Y _____ N _____

Score: 0 1 2 3

Appendix D

Homonym Judgment Task (Doherty, 2000)

Now we are going to play a pointing game! I am going to say a word and then I want you to point the picture that matches the word.

Phase 1: Vocabulary Check

Set A (Bat) - Distractor items: dog, flower

Can you point to the bat?

After identifies the bat (flying) cover the picture with the sticky note.

Which one of these is the bat?

*If the child points to the bat under the sticky note say: **which of THESE THREE is the bat?***

*Uncover the all pictures. **So this is a bat and this is a bat! See, they have the same name!***

Set A ____/1

Set B (letter) - Distractor items: bus, apple

Can you point to the letter?

After identifies the letter (alphabet) cover the picture with the sticky note.

Which one of these is the letter?

*If the child points to the letter under the sticky note say: **which of THESE THREE is the letter?***

*Uncover the all pictures. **So this is a letter and this is a letter! Cool, they have the same name!***

Set B ____/1

Set C (nail) - Distractor items: cat, blocks

Can you point to the nail?

After identifies the nail cover the picture with the sticky note.

Which one of these is the nail?

*If the child points to the nail under the sticky note say: **which of THESE THREE is the nail?***

*Uncover the all pictures. **So this is a nail and this is a nail! Wow, they have the same name!***

Set C ____/1

Set D (Glasses) - Distractor items: bus, apple

Can you point to the glasses?

After identifies the glasses cover the picture with the sticky note.

Which one of these are the glasses?

*If the child points to the glasses under the sticky note say: **which of THESE THREE are the glasses?***

*Uncover the all pictures. **So these are glasses and these are glasses! See, they have the same name!***

Set D ____/1

Set E (Night/Knight) - Distractor items: cow, book

Can you point to the night?

After identifies the night cover the picture with the sticky note.

Which one of these is the knight?

*If the child points to the night under the sticky note say: **which of THESE THREE is the night?***

*Uncover the all pictures. **So this is a night and this is a knight! See, they have the same name!***

Set E ____/1

Score: ____/ 5

*****Must know at least 4/5 homonym pairs to continue*****

Phase 2: Modelling

(Put on puppet) This is my friend, Donald Duck, and he is going to play the game with us. Ready?

Can you point to the 'letter'? Circle: Alphabet Mail
If chooses both say – Please pick just one!

Now in this game, it's Donald's job to point to the other letter, not the same one you pointed to. Let's practice. Remember you pointed to this letter (*point*). So which letter should Donald point to?

Donald should point to this letter (*point – OTHER LETTER*) because that is the other letter, not the one you pointed to!

So, Donald should NOT point to this letter (*point - SAME LETTER*) because that's what you pointed to.

And, Donald should NOT point to the bus (*point – BUS*) because it is his job to point to the other letter.

Phase 3: Test

*****Note - If child did not get one of these homonym pairs in vocab. phase, replace with NAIL*****

Okay – Let's try some more! Remember it is Donald duck's job to point to the other picture that has the same name.

1. Can you point to the bat?

Now it's Donald's turn.

Donald points to the OTHER BAT – Is that what Donald Duck's should've done?

Y _____ N _____

2. Can you point to the night?

Now it's Donald's turn.

Donald points to the BOOK - Is that what Donald Duck should've done?

Y _____ N _____

3. Can you point to the glasses?

Now it's Donald's turn.

Donald points to the OTHER GLASSES – Is that what Donald Duck should've done?

Y _____ N _____

Score: 0 1 2 3

Appendix E

1. **Snow/Grass** (Carlson & Moses, 2001)

Materials: Green square, white square, felt hand prints for children to place their hand

Notes: 1. Trial is considered self-corrected if child originally moves towards the wrong colour, but then changes directions and chooses the right colour. 2. Timing through the trials should be paced appropriately, leaving enough time for children to return to the red handprints each time. 3. Provide reminders if child is not returning to handprints.

Can you tell me what is the color of snow?

Can you tell me what is the color of grass? Very good! (*Place green and white squares on table in front of child approximately 12-18 inches apart from one another*)

Now, we are going to play a silly opposites game, where every time I say snow I want you to point to the green color (point to green square) and every time I say grass, I want you to point to the white color (point to white square).

Let's practice, first please put your hands on the red hand shapes on the table (*Tape handprints to table*). **Please put both your hands back on the handprints each time. Ready?**

Grass. Snow. These are teaching trials, and instructions can be repeated if child does not understand.

Trials will be administered in the following fixed order:

Target Card	Response	Score
1. Grass	White	2 1 0
2. Snow	Green	2 1 0
3. Snow	Green	2 1 0
4. Grass	White	2 1 0
5. Snow	Green	2 1 0
6. Grass	White	2 1 0
7. Grass	White	2 1 0
8. Snow	Green	2 1 0
9. Snow	Green	2 1 0
10. Grass	White	2 1 0
11. Snow	Green	2 1 0
12. Grass	White	2 1 0
13. Grass	White	2 1 0
14. Snow	Green	2 1 0
15. Grass	White	2 1 0
16. Snow	Green	2 1 0

Total Score Grass/Snow: _____ / 32

Scoring: 0 for pointing incorrectly, 1 for self-correcting, 2 for correct pointing.

Appendix F

HEADS-TOES-KNEES-SHOULDERS

Note:

- Scoring:
 - 2 = child produces the correct response immediately
 - 1 = self-correct right away, without prompting
 - 0 = If they do not touch the correct part of their body at all
- Self-correction: Mark “self-correct” on both the training and testing portion if the child makes any discernible motion toward the incorrect answer, but then changes his/her mind and makes the correct response. Pausing to think, not moving, and then responding correctly does not count as a self-correction.
- Task is done standing

PART 1

Have you ever played the game “Heads Shoulders Knees and Toes?” Well this game is kind of like that one, but it’s a silly game.

If I said, “touch your head” what would you normally do?

That’s right, I would touch my head! [*demo this*]

And if I said “touch your toes” what would you normally do?

That’s right, I would touch my toes! [*demo this*]

But this is going to be a silly game, so you’re going to do the opposite of what I say! So when I say “touch your head,” I want you to touch your toes [*demo this*] and when I say “touch your toes” I want you to touch your head! [*demo this*] Good Job!

Part 1 Training:

				Retraining
A1. What do you do if I say “touch your head”?				
0 (head)	1	2 (toes)		_____
A2. What do you do if I say “touch your toes”?				
0 (toes)	1	2 (head)		_____

Okay, let’s practice!

Part 1 Practice:

You can correct the kids during practice trials if they make any mistakes. Say things like “Remember, we’re playing a silly game so you have to do the opposite of what I say”

	Incorrect	Self-Correct*	Correct	
B1. Touch your head	0 (head)	1	2 (toes)	Retraining _____
B2. Touch your toes	0 (toes)	1	2 (head)	
B3. Touch your head	0 (head)	1	2 (toes)	
B4. Touch your toes	0 (toes)	1	2 (head)	

Note: Retraining occurs only 3 times

Alright, ready for the real game? Remember, we're playing the silly game!

Part 1 Testing:

-Do not correct child. If child gets a score > 10, continue to next part, otherwise discontinue.

		Incorrect	Self-Correct*	Correct
21.	Touch your head	0 (head)	1	2 (toes)
22.	Touch your toes	0 (toes)	1	2 (head)
23.	Touch your toes	0 (toes)	1	2 (head)
24.	Touch your head	0 (head)	1	2 (toes)
25.	Touch your toes	0 (toes)	1	2 (head)
26.	Touch your head	0 (head)	1	2 (toes)
27.	Touch your head	0 (head)	1	2 (toes)
28.	Touch your toes	0 (toes)	1	2 (head)
29.	Touch your head	0 (head)	1	2 (toes)
30.	Touch your toes	0 (toes)	1	2 (head)

Total Points: _____

Number of I responses: _____

PART 2

Okay, because you did such a great job, we're going to add 2 new actions! When I say "touch your shoulders" I want you to touch your knees [demo] and when I say "touch your knees" I want you to touch your shoulders! [demo]

<p>C1. What do you do if I say "touch your knees?"</p> <p>0 (knees) 1 2 (shoulders)</p>	<p>Retraining</p> <p>_____</p>
---	---------------------------------------

Okay, let's practice!

Part 2 Practice:

You can correct the kids during practice trials if they make any mistakes. Say things like "Remember, we're playing a silly game so you have to do the opposite of what I say"

	Incorrect	Self-Correct*	Correct	
D1. Touch your knees	0 (knees)	1	2 (shoulders)	Retraining _____
D2. Touch your shoulders	0 (shoulders)	1	2 (knees)	
D3. Touch your knees	0 (knees)	1	2 (shoulders)	
D4. Touch your shoulders	0 (shoulders)	1	2 (knees)	

Good job! Now we're going to combine all 4 actions! So I might say "touch your head," "touch your toes," "touch your knees" or "touch your shoulder." Remember, we're playing the silly game so you have to do the opposite of what I say, okay?

Practice 2 Testing:

	Incorrect	Self-Correct	Correct
31. Touch your head	0	1	2 (toes)
32. Touch your toes	0	1	2 (head)
33. Touch your knees	0	1	2 (shoulders)
34. Touch your toes	0	1	2 (head)
35. Touch your shoulders	0	1	2 (knees)
36. Touch your head	0	1	2 (toes)
37. Touch your knees	0	1	2 (shoulders)
38. Touch your knees	0	1	2 (shoulders)
39. Touch your shoulders	0	1	2 (knees)
40. Touch your toes	0	1	2 (head)

Total Points: _____

Number of I responses: _____

Appendix G

False Sign (location) (Sabbagh, Moses & Shiverick, 2006)

Okay, I have a new game to show you.

Arrow starts pointing up.

Here is a board and this is an arrow (point to the arrow). There is a green house (point to green house) and a red house (point to red house). Kim likes to use this arrow to let people know which house she is playing in. Let's practice.

(Point arrow to green house – then point arrow up)

Where does the arrow say Kim is? GREEN RED

(Point arrow to red house– then point arrow up)

Where does the arrow say Kim is? GREEN RED

(Point arrow to red house– then point arrow up)

Where does the arrow say Kim is? GREEN RED

(Point arrow to green house– then point arrow up)

Where does the arrow say Kim is? GREEN RED

This is Kim (hold up Kim), and this is her friend Sam (hold up Sam) and they are trying to decide which house to play in: the red house or the green house.

But, then Sam says he is hungry and is going to get a snack. Before he goes he asks Kim: How will I know where to find you when I get back?

Kim says, I'll leave the arrow pointing to where I am so you can find me.

Then, Sam leaves to get a snack (put Sam under table).

Kim changes arrow and goes into red house. Then, Kim goes to play in the red house.

After a while, Kim decides she wants to play in the green house instead (move Kim to green house). But-- Silly Kim! She forgot to change the arrow!

Then, Sam comes back and looks for Kim.

False Sign Test Question: **Where does the arrow say Kim is?** Response: _____

Reality Control Question: **Where is Kim really?** Response: _____

Score: 0 1

***Note – must answer both questions correctly to receive a score of 1.

Appendix H

False Sign (contents) (Ioa, Leekman, Perner & McConachie, 2011)

Check this out! (*show child the box*) – *Crayon box*

What do you think is inside this box?

Open box and show child contents.

Look there are (*request that the participant say what's inside. If he/she doesn't tell them and ask them to repeat*)

Return box to its original state.

Now what do you think is in the box?

Circle: CANDLES CRAYONS OTHER: _____

What does the label on the box show what's inside?

Circle: CANDLES CRAYONS OTHER: _____

Score: 0 1

*Note – must answer both questions correctly to receive a score of 1.

Appendix I

Colour-Colour Control (Perner, Stummer, Sprung & Doherty, 2002)

Phase 1: Vocabulary check

Can you show me the blue square / yellow/ green / red? (Next sheet) Purple / grey / orange / pink.

Phase 2: Modelling

(Put on puppet) **This is my friend, Mickey Mouse, and Mickey is going to play the game with us. Okay?**

Here is a TREE. This tree has two colours. Can you tell me ONE colour you see on this TREE? Good job.

Now in this game, it's Mickey's job says the other colour, not the one that you said. Let's practice. Remember you said *colour*. So, which colour should Mickey say? *If correct say - right!*

Mickey should say (OTHER COLOUR) because that is the other colour of this tree!

Mickey should NOT say (SAME COLOUR) because that it what you said.

And, Mickey should NOT say (PINK) because this tree is not pink.

Phase 3: Test

Okay – Let's try some more! Remember it is Mickey's job to say the other colour, not the colour you said.

1. Can you tell me one colour that you see on this SUN? Circle: YELLOW ORANGE

Mickey says: OTHER COLOUR

Did Mickey say what he was supposed to say?

Y _____ N _____

2. Can you tell me one color you see on these SOCKS? Circle: ORANGE BLUE

Mickey says: PURPLE

Did Mickey say what he was supposed to say?

Y _____ N _____

3. Can you tell me one colour that you see on this FLOWER? Circle: RED BLUE

Mickey mouse says: OTHER COLOUR

Did Mickey say what he was supposed to say?

Y _____ N _____

Score: 0 1 2 3