A Comparison of False Belief and Referential Opacity Tasks: Sorting out the Relations

By

Dawn Mullins B.A. (Hon.)

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Arts

Department of Psychology

Carleton University

Ottawa, Ontario

2004 May 7

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-93916-2

Our file Notre référence

ISBN: 0-612-93916-2

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this dissertation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de ce manuscrit.

While these forms may be included in the document page count, their removal does not represent any loss of content from the dissertation.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The purpose of the present research was to investigate relations among tasks that measure children's metarepresentational ability. False belief tasks were compared to three different types of referential opacity tasks within the same group of children. Sixty children aged 3;9 to 6;11 years received 3 false belief and 9 opacity tasks. Children performed better on false belief than on opacity tasks. Performance on false belief tasks was related to one type of opacity task. Two of the three opacity tasks were related to each other; the third was unrelated to either of the other tasks. False belief tasks measured an early developing metarepresentational ability. The opaque contexts measured a more sophisticated metarepresentational ability not yet mastered by the children in the present study. Differences in the way co-referential terms were presented in opacity questions on the different types of tasks may have differentially supported the children's metarepresentational ability.

Acknowledgements

I would like to thank Deepthi Kamawar for her guidance, patience, and outstanding support for the past year and one-half; Jo-Anne LeFevre for her insight, advice and direction for the past three years; John Logan for answering my questions no matter how trivial; Craig Leth-Steensen for his statistics advice; Laura Grace Conlon for her charming and delightful illustrations; Etelle Bourassa for keeping me “administratively correct”; my husband Jim and children Stephen and Amy for their love, encouragement, and humour.

To all of the people who helped with this study, I am grateful for their time, enthusiasm and cooperation: the parents, students and staff of Kanata Research Park Family Centre, The Children’s Place, Carleton Memorial Day Care Inc., Churchill-Carling Daycare, St. Thomas Elementary School, Castlefrank Elementary School, and Katimavik Elementary School. Also, thank-you to Dr. William Tomka for his enthusiasm for the project as a member of the Ottawa-Carleton Research Advisory Committee and for his letter of endorsement to parents as principal of St. Thomas School. Finally, a special thanks to a bright and lovely six and one-half year old girl who did an amazing job of replicating my tasks with her toys, creating a story and questions, administering the tasks to me, recording my responses, and doing me the honour of saying she wants to do what I do when she grows up. It doesn’t get any better than that!

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	v
List of Figures	vii
List of Appendices	viii
Introduction.....	1
Method.....	35
Results.....	43
Discussion.....	76
References	96

List of Tables

Table 1.	Summary of scores for false belief tasks.	38
Table 2.	Summary of scores for opacity tasks.	40
Table 3.	Summary of tasks in order one and order two.	43
Table 4.	Performance on false belief and DF tasks.	46
Table 5.	Numbers (%) of children who answered questions correctly on AR tasks.	49
Table 6.	Performance on false belief and AR A opacity tasks.	50
Table 7.	Performance on false belief and KO opacity tasks.	52
Table 8.	Mean performances and standard deviations on false belief vs. KO opacity tasks per age group.	54
Table 9.	Mean performances and standard deviations on false belief vs. KO opacity tasks per age group.	56
Table 10.	Significant age effects across false belief and opacity tasks.	60
Table 11.	Previous and present findings for false belief and opacity tasks.	61
Table 12.	Mean performances and standard deviations on age, false belief, digit span backwards, for gender and order.	64
Table 13.	Mean performances and standard deviations on opacity tasks for order one and order two.	66

Table 14. Means and standard deviations for opacity by age and order.	67
Table 15. Partial correlation matrix: age partialled out.	70
Table 16. Partial correlation matrix: age partialled out for AR A and AR B.....	70
Table 17. Raw correlations among all the tasks.	72
Table 18. Summary of hierarchical regressions analyses predicting KO referential opacity. ..	74
Table 19. Summary of hierarchical regressions analyses predicting DF referential opacity.	75
Table 20. Summary of hierarchical regressions analyses predicting AR A referential opacity.	75

List of Figures

Figure 1. Mean scores for false belief tasks and each type of opacity tasks across age groups.	57
Figure 2. Mean opacity score as a function of opacity task, gender, and order.	62
Figure 3. Mean opacity scores as a function of opacity task, order, and age.	65

List of Appendices

Appendix A Belief tasks	99
Appendix B Opacity tasks	101
Appendix C Complementation task	114
Appendix D Digit span tasks	115
Appendix E Consent form	117
Appendix F Design of experiment	119
Appendix G Comparisons of critical questions across opacity tasks	120

Introduction

Children's understanding of the mind develops gradually over the preschool years. The mind is made up of mental states or mental representations such as beliefs, desires, emotions and intentions (Astington, 1993). Remarkably, children as young as 3 years old understand that other people have desires and beliefs that govern their actions (Wellman, 1990, as cited in Siegler, 1993). This commonsense or naive understanding of how the mind works is called 'theory of mind'. As cited in Astington (1993), the term theory of mind (ToM) was initially used by Premack and Woodruff (1978) in research on primate intelligence. Specifically, they examined chimpanzees' ability to predict human action and concluded that:

An individual has a theory of mind if he imputes mental states to himself and others. A system of inferences of this kind is properly viewed as a theory because such states are not directly observable, and the system can be used to make predictions about the behaviour of others (p. 4).

Not surprisingly, Premack and Woodruff's research sparked much lively debate among developmental psychologists; if primates could be said to have a theory of mind, why not children? The concept of the mind, like the concepts of time, space, and number, is considered to be one of a few central concepts of special importance to the developing child. These core developmental concepts are important "because they are used to represent a vast range of experiences, because they are present in some form from infancy to old age in all of the world's cultures, and because understanding the world would be impossible without them" (Siegler, 1993, p. 245). The goal of the present research is to further our knowledge about how children come to understand mental events by examining young children's performance on tasks that investigate and measure the nature of children's abilities to represent other people's beliefs and knowledge.

Children's understanding of the concept of mind develops early. By the end of infancy, children can represent situations according to time (past, present or future), pretence (real or counterfactual), and can distinguish real people and objects from those in mirrors and pictures (Perner, 1992). Bretherton, Mcnew and Beeghly-Smith (1981) noted that the first evidence for theory of mind in humans emerges at the end of the first year with language acquisition when children engage in joint attention behaviours such as pointing and naming of objects (as cited by Astington & Jenkins, 1999).

By the age of two, toddlers show concern about other people's emotional reactions, and by age three, children understand that people who get what they want are happier than people who don't. Expressions of negative emotions such as sadness or anger by other people are commonplace occurrences. Hence, understanding or 'reading' other people's minds is important for successful social interaction and is adaptive for the young child.

Toddlers are able to form theories about how people behave in specific situations; that is, they are situation theorists (Perner, 1992). In any given situation, for children to be able to ascribe mental states to themselves and others they must be able to represent the situation, represent how the other person represents the situation, and represent any differences in the way s(he) represents the situation compared to the other person (Perner, 1992). According to Perner, the situation theorist considers other people's knowledge as it relates to a real situation but does not understand the mind as a representational system.

Children must grasp the concept of representation to progress from situation theory to a representational theory of mind. Specifically, children must understand the properties of representation: 1) representations exist and represent something; 2) what is being

represented (representational content) can be real or unreal; and 3) representations represent something (referent) as being a certain way. Hence, misrepresentation is possible (Perner, 1992).

Children are credited with a representational theory of mind when they are able to represent other people's representations of the world; that is, when they have metarepresentational ability. However, other people's representations of the world may or may not depict reality. Thus, when considering another person's representation of the world, children must be able to predict the other person's behaviour even when that person's representation is incomplete or inaccurate (Kamawar, 2000). Children must come to understand that people act on their representations of the world even if their mind misrepresents reality (Astington, 2001). As a consequence of understanding that other people act on their representations, children's interactions with others are altered; for example, children begin to understand deception and competitive games (Perner, 1992). Thus theory of mind development is important to children's general social-cognitive development. For, example, understanding that other people behave according to their representations or perspectives of the world is crucial for children's successful social interaction. Children's capacities to empathize, comfort, frustrate, tease, deceive, negotiate, and compromise in conflict are all linked to their understanding that people behave according to their individual beliefs and perspectives (Dunn & Cutting, 1999).

Numerous tasks have been developed to investigate and measure the nature of children's abilities to represent other peoples' beliefs and knowledge. Two categories of tasks, false belief and opaque contexts are the focus of this study.

False Belief Tasks

The first type of false belief task, known as the “change of location task”, was introduced by Wimmer and Perner in 1983. They argued that the false belief paradigm is a means of measuring children’s metarepresentational ability. This paradigm involves a complex metarepresentational problem to test whether children can explicitly represent the difference between their own and someone else's representation of the same situation. They argued that if children accurately represent this difference they should be able to predict the other person’s behaviour. A typical false belief task involves telling children a story using characters that have a false belief because they have outdated and incorrect knowledge about a situation. The child participants, who have accurate and up-to-date knowledge of the situation, answer questions about how the character will respond based upon the character’s false belief or incorrect knowledge. For example, in the change of location or unseen displacement task a child is told the following story, acted out in front of her (Wimmer & Perner, 1983):

Mother returns from her shopping trip. She bought chocolate for a cake. Maxi may help put away the things. He asks her: “Where should I put the chocolate?” “In the blue cupboard”, says the mother. “Wait I’ll lift you up there, because you are too small.” Mother lifts him up. Maxi puts the chocolate into the blue cupboard. (A toy chocolate is put into the blue match box). Maxi remembers exactly where he put the chocolate so that he could come back and get some later. He loves chocolate. Then he leaves for the playground. (The boy doll is removed). Mother starts to prepare the cake and takes the chocolate out of the blue cupboard. She grates a bit into the dough and then she does *not* put it back into the *blue* but into the *green* cupboard. (Toy chocolate is thereby transferred from the blue to the green matchbox). Now she realizes that she forgot to buy eggs. So she goes to her neighbour for some eggs. There comes Maxi back from the playground, hungry, and he wants to get some chocolate. (Boy doll reappears). He still remembers where he had put the chocolate.” False belief question: Where will Maxi look for the chocolate? (p. 109).

The correct answer is that Maxi will look in the blue cupboard (i.e., where Maxi had first placed it).

According to Wimmer and Perner (1983), representing false beliefs requires:

(1) the construction of two different models of the world (the false belief and the actual situation); and (2) an explicit representation of the false relation between the propositions in one model and the corresponding propositions in the other model. That is, children must understand that their own accurate representation of reality differs from the other person's inaccurate representation of reality, and that the other person will act on their misrepresentation of reality. In other words, participating children must represent the story character's representation in order to correctly predict the character's actions (i.e., metarepresentation is necessary). Wimmer and Perner found that children aged 4-5 are fairly successful on this task whereas, children aged 3 are not. Thus, according to these researchers, from about age four onwards children are thought to have developed the ability to metarepresent.

Perner, Leekam and Wimmer (1987) developed another false belief task (the 'deceptive box' or 'unexpected contents' task) that is widely used and has also become a standard false belief task (as cited in Astington 1993): children are shown a Smarties box and asked what they think is in the box. Generally, they say "Smarties". Then the box is opened and they are shown that the box really contains a pencil. The pencil is put back inside and the box is closed. The children are then asked what was in the box (the reality question). Next, the children are asked what they had thought was in the box before it was opened (a false belief question for "self") and what someone else, who hadn't seen inside the box would think was in it (a false belief question for "other"). The correct answer for both questions was 'Smarties' (or candy, or whatever the child had said to the first question etc.).

To succeed on this task children need to distinguish between two different types of objects (Smarties vs. pencils), represent their initial representation of the contents of the box and that it differed from what was actually in the box, and represent this situation from another's perspective. Hence, similar to the change of location task, children must understand that their own accurate representation of reality (contents are pencils) differs from the other person's inaccurate representation of reality (contents are Smarties), and that the other person will act on their misrepresentation of reality.

Controversy surrounding Wimmer and Perner's 1983 and Perner et al.'s 1987 tasks has stimulated the development of a large body of research over the past twenty years. Because false belief tasks have been studied extensively, they are considered diagnostic of the "mentalistic understanding of persons" (Wellman, Cross & Watson, 2001, p. 656). For example, false belief tasks are an important measure in the study of individual differences in young children's social cognition. In this area of research, the tasks are used to assess the influence of early family conversations, engagement in pretend play, family structure, and development of mentalistic understandings. False belief tasks are also a major tool for research with developmentally delayed individuals. Specifically, high- functioning autistic people fail false belief tasks, whereas other delayed populations (e.g. Down's syndrome) of equivalent mental age do not. Researchers propose that autistic people are unable to acquire a theory of mind because severe social disconnectedness prevents them from considering the mental representations of other people. In contrast, these autistic individuals can reason competently about physical phenomena (Wellman et. al, 2001).

The distinction between understanding physical and mental phenomena is important because it distinguishes false belief tasks from other developmental tasks that examine reasoning or perspective-taking based on physical and not mental events. For example, in Piaget's three mountain task children are asked to consider the perspective of others by identifying aspects of a physical model as viewed by a doll placed in different positions. The different perspectives are all true, they are just considered from different vantage points. In contrast, false belief tasks examine children's understanding of other people's beliefs. For example, if one person believes that the chocolate is in location A, but, the child knows it is really in location B, then these two perspectives are not both true. The child must understand that the other person's perspective or belief is incorrect or false. Thus, false belief tasks involve truth-incompatible perspectives that contain a criterion for evaluating the truth or falsity of the situation (i.e. where the chocolate really is). In contrast, tasks like the three mountain task involve truth-compatible perspectives that contain different perspectives on one true situation (Perner, Stummer, Sprung & Doherty, 2002).

There are other important differences between current theory of mind research and perspective taking research in the 1960s and 1970s: different definitions of perspective-taking were used and the tasks and measurement tools used in perspective-taking research varied greatly; theory of mind research uses standard false belief tasks defined in terms of metarepresentational development. Furthermore, "perspective-taking research assumed young children were egocentric and could not understand another person's viewpoint, whereas theory-of-mind research integrates children's understanding of their own and other people's mental states" (Astington & Jenkins, 1999,

p.12). Hence, a fundamental shift in researchers' thinking about how children understand other people's minds has led to theory of mind research.

Similar to Piagetian tasks, false belief tasks are central to current social-cognitive developmental research, and initial accounts of false belief tasks and claims of conceptual change have all been vigorously challenged. One of the most controversial points is the claim that a radical conceptual shift takes place around 4 years of age (and not before) whereby children acquire a fully representational theory of mind comparable to that of adults (Chandler & Sokol, 1999). Many researchers have tried to find evidence of metarepresentational ability before the pivotal age of 4 years. DeVilliers and Pyers (2002) noted that some aspects of theory of mind are precursors to false belief; they cited the findings that young children read intentions from eye gaze and pointing (Tomasello, 1995), can understand other people's desires (Gopnik, 1993), know the difference between pretence and reality (Leslie, 1987), and use mental state verbs such as "want", "think" and "know" (Wellman, 1990). These activities are generally considered to play a vital role in supporting and fostering children's developing awareness of mental states (Hughes, 1998). In a recent meta-analysis of false belief tasks, Wellman, Cross and Watson (2001) note that false belief tasks have been criticized for masking conceptual competence in young children because the tasks are too demanding or confusing. They suggest that two important factors influence performance on any cognitive task: competence or conceptual understanding of the problem and cognitive skills such as the ability to attend to the information, remember relevant information, comprehend and express answers to questions. Considering the required skills, researchers have designed

many versions of the task over the years to enhance children's performance by reducing task demands.

Wellman et al. included 591 false-belief conditions in their meta-analysis and drew several important conclusions: 1) an important conceptual change in children's understanding of other people's minds takes place between 2;6 and 5 years of age; 2) this change is evident across many cultures, non-schooled as well as literate; 3) many variations in stimuli and procedures across numerous tasks were irrelevant to successful performance; 4) reducing task demands does improve performance, but does not change the developmental trajectory; 5) even with the effects due to age, task demands, and information processing limitations controlled, the developmental pattern was still significant.

One strong recommendation by Wellman et al. (2001) arising from their meta-analysis was to compare false belief tasks to other well-chosen tasks to promote better understanding of children's theory of mind development. Indeed, some researchers (Apperly & Robinson, 1998; Kamawar & Olson, 1999) argued for continued metarepresentational development beyond the age of four and suggest a more gradual development rather than a radical shift towards adult metarepresentational competence. These researchers have developed tasks in the area of opaque contexts to test for more complex metarepresentational understanding. Hence, comparison of false belief to opacity tasks can make an important contribution to the ToM literature and inform us about children's cognitive development.

Opacity Tasks Measure Sensitivity to Opaque Contexts

To explain what is meant by referentially opaque contexts, it is useful to first explain non-opaque or transparent contexts. Kamawar (2000) uses the well-known example by Frege (1892/1991) to explain transparent and opaque contexts: in transparent contexts, substitution of a co-referential term (a term that refers to the same object) does not change the truth value of a sentence. For example, Venus can be described as both the "Morning Star" and the "Evening Star". Thus, given that "Venus is the Morning Star" is true, then substituting "Evening Star" will not change the truth of the resulting sentence because these are co-referential descriptions. In contrast, substitution of a co-referential term in opaque contexts *may* alter the truth value of a sentence. For example, if we start with "John knows that Venus is the Morning Star" is true and then make the same substitution we get "John knows that Venus is the Evening Star". However, this will be false if John only knows Venus as the Morning Star. John is only partially informed and knows one description but not the other. The person whose mental state is described (i.e., John) may not know that Venus is also referred to as the Evening Star. Hence, he does not have a mental representation of Venus *as* the Evening Star; therefore, it cannot be said of him that he "knows that Venus is the Evening Star". Like false belief tasks, the transparency/opacity distinction requires that a child must learn to understand /represent /predict other people's behaviour in terms of their representations and mental states (Kamawar, 2000). Thus, there is a theoretical link between the two categories of tasks.

In contrast to false belief tasks, opacity tasks have not been studied extensively. Different kinds of opacity tasks involving direct quotations, indirect quotations, intentional verbs (e.g., meant to), and co-referential terms to pick out the same referent have recently been employed to test children's metarepresentational ability. However,

only referential opacity tasks (using co-referential terms) will be considered here. Examples of transparent and referential opacity questions are found in Apperly and Robinson (2002). Children are shown an object that can be known under two descriptions. For example, a ball is also a present but, only one description is apparent from looking at the object (i.e., “ball”). Next, Heinz the puppet is introduced. Heinz is allowed to see the dual identity item in a box but, he is not told that it is a present. Children are asked a transparent (substitution-insensitive) question: “Can Heinz see the present inside the box?” and two questions about what Heinz knew: A hard opaque (substitution-sensitive) question “Does Heinz know there’s a present in the box?” was given to one group of children and an easier opaque question was given to a different group of children “Does Heinz know that the ball is a present?” Researchers in the area have employed a variety of opacity tasks to measure metarepresentational ability in opaque contexts and compare children’s performance on opacity and false belief tasks. Specific tasks used by different researchers will be discussed next. All of the false belief tasks used by the researchers are modeled after the standard unexpected contents and change of location tasks described earlier. Therefore, only brief descriptions of these tasks will be given. In contrast, the opacity tasks used by the researchers differ considerably. Thus, a detailed example of each kind of opacity task will be given.

Empirical Work

DeVilliers and Fitneva (1996) conducted the first study to compare children’s performance on false belief and referential opacity tasks. They drew upon earlier work by Russell (1987) who found that children as old as 5;6 years had difficulty with referential opacity tasks. However, Russell’s questions to the children were very complex (e.g. “Can

we say that George was thinking, I must find the man with the curly red hair who stole my watch?”). Furthermore, Russell did not compare false belief tasks to referential opacity tasks. DeVilliers and Fitneva simplified the questions for the children, extended Russell’s work by comparing false belief and referential opacity tasks, and included younger children. Specifically, they compared two false belief tasks to two versions of a referential opacity task in children aged 3-6 years. In addition, these researchers maintained that for children to understand other people’s mental states, it is necessary for them to understand the language structure that refers to mental states. Mental and communication verbs have special grammatical status in that they allow for representations in the mind that differ from reality (e.g., “What did she say she bought?” “She said she bought paper towels but she bought cake.”). The researchers propose that children can understand another person saying something false before they can understand that another person can believe something is false.

They were particularly interested in examining how referential substitution with verbs of communication compared to mental and action verbs. Examples of the three types of verbs used are: an action verb (put), a mental verb (know) and two communication verbs (ask and say). The action verb “put” measures children’s performance in the transparent context, the mental verb “know” is used in the opaque context, and the two communication verbs, “ask” and “say” are used to examine children’s handling of a specific aspect of complex syntax (tensed complements). An example of a tensed complement under a communication verb was given above, “What did she say she bought?”

While both mental and communication verbs share the same syntactic structure, deVilliers and Fitneva used children's ability to handle tensed complements with communication verbs as their measure of complementation ability. The reason for doing so was to ensure that a correlation between handling tensed complements and false belief (mental state understanding) wasn't confounded by using mental verbs in the complementation task (where it could be argued that mental verb complementation is just related to handling mental verbs). It is worth noting that opaque contexts are examples of tensed complement structures.

The false belief stories were:

1) An unexpected contents task using a Crayola box that contained a toy plastic fork. There was a question for 'self' and a question for 'other': the children were asked what they had thought was in the box before it was opened and what a friend, who hadn't seen inside the box, would think was in it (not always in this order). The correct answer for both questions was 'crayons'. The child got one point for correctly remembering his/her own false belief and another point for correctly predicting his friend's false belief.

2) A change of location task in which a story character puts a cake in a cupboard, and while he is out playing, his Dad moves the cake to the refrigerator. There are two check questions: Where did Bobby put the cake? Where is it now? There is one false belief question: When he comes in the kitchen, where will Bobby look for the cake? The child is given one point for passing this task: the child must correctly answer the memory question and predict that Bobby will look in the cupboard. A child would fail if he passed the memory question but said Bobby would look in the refrigerator. Thus, children were only credited as

passing false belief tasks if they scored a total of three points across the two false belief tasks.

The opacity story was:

This silver box is a birthday present. This birthday present is candy.
The Mom put the silver box on the top shelf.

Did the mom put the silver box on the top shelf?
Did the mom put the birthday present on the top shelf?

This little girl walked into the room one day and saw a silver box on the top shelf. She thought, "I wonder what that silver box is?"

Does the little girl know the silver box is on the top shelf?
Does the little girl know the candy is on the top shelf?

She went to her Mom and said, "Mom, why is there a silver box on the shelf!"

Did the little girl ask why the silver box was on the shelf?
Did the little girl ask why the birthday present was on the shelf?

Her Mom said, "Do not touch the silver box!"

Did the mom say that the girl should not touch the silver box?
Did the mom say that the girl should not touch the candy?
But the girl took the silver box and put it on the table.

Did the girl put the silver box on the table?
Did the girl put the birthday present on the table?

The Mom came in and saw the silver box on the table.

Does the mom know the silver box is on the table?
Does the mom know the candy is on the table? (p. 11).

The child participant and one of the story characters has full knowledge of the situation (Mom) whereas, the other story character (little girl) has only partial knowledge. The container is opaque therefore the object referent has one visible identity (silver box)

based on appearance, one invisible identity (candy) based on the contents, and one invisible identity (birthday present) based on the function.

Two versions of the stories were used; in one version questions referred to contents (candy) whereas in the other version questions referred to function (birthday present). For each age group, half of the children received one version and half the other version. Thirteen adults also served as controls on the opacity tasks to ensure that the answers considered “correct” were in fact the responses that adults would have given. Their performance was almost perfect; 95% of the adults answered correctly. Each participant was given a score of 1 for each “yes” answer for a possible total of 8; 4 points for the standard reference (e.g. silver box) and 4 points for the substitution (e.g. candy). A difference score between these answers was derived to measure whether referential substitution was permitted or not. The scores ranged from -4 to +4 with 0 as the best score for transparency and +4 the maximum for correctly not using substitution in an opaque context, -4 was the worst score for the unlikely event of preference for the substituted reference over the original one (i.e., a type of reverse opacity).

DeVilliers and Fitneva (1996) found that children who passed the false belief tasks (mean age of 5 years) also passed the opacity tasks and those who failed on the false belief tasks (mean age of 3:8 years) also failed the opacity tasks. As predicted, each of the three types of verbs differed significantly from the other two. Children who passed false belief tasks correctly refrained from substitution for mental verbs with the unknowing (ignorant) protagonist. Furthermore, children who passed false belief tasks treated verbs of communication as intermediate between action and mental verbs. That is, they substituted in transparent contexts and refrained from substitution in opaque

contexts generally although, not as well as adults. In contrast, children who failed false belief tasks were unable to determine when substitution of a co-referential term was or was not allowed; they did not distinguish among the three types of verbs. However, their results were no longer significant once age was controlled.

DeVilliers and Fitneva (1996) concluded that passing false belief tasks is a prerequisite but may not be sufficient for performance on opacity tasks. They suggest that above and beyond needing to pass false belief tasks, children need to master the syntax of tensed complements. To further explore when children master the difference between referential substitution for communication compared to mental verbs, another study was done in 2002 by DeVilliers and Pyers.

DeVilliers and Pyers (2002) conducted a longitudinal study comparing false belief understanding to measures of language production and comprehension over the course of one year. The children were between 3 and 5 years old at the beginning of the study. The researchers predicted that the particular syntactic ability of complementation was the most important ability required for success on both false belief and referential opacity tasks. Specifically, these researchers note that mental verbs (e.g., think) and communication (e.g., say) have special grammatical status in that they allow for representations in the mind that differ from reality (e.g., "he thought he found his ring, but it was really a bottle cap"). They found that mastering complements predicted false belief performance but not vice versa. Hence, DeVilliers and Pyers suggest that mastery of embedded complements is a precursor and possibly a prerequisite for successful performance on false belief and opacity tasks (which share syntactic structures).

Apperly and Robinson (1998) conducted two experiments to examine differences in difficulty between false belief and referential opacity tasks. They set out to challenge Perner's (1991) claim that children around four years of age have acquired a complete understanding of the mind because they are able to pass false belief tasks. If children have a complete understanding of the mind, then, they should have no difficulty with referential opacity tasks. In contrast to this prediction, Russell (1987) provided evidence for children's lack of understanding of the mind beyond age four years when children in his study could not pass his opacity tasks. Russell's work was criticized because his tasks were too complex. Like deVilliers and Fitneva, Apperly and Robinson drew upon Russell's research to create their own tasks but simplified the questions.

To minimize difficulty on the opacity tasks, Apperly and Robinson used only the mental verb "know" and reduced the number of embedded clauses to one. They created questions that were similar in form to those used in theory of mind research. Recall that children were asked a transparent (substitution-insensitive) question: "Can Heinz see the present inside the box?" and two questions about what Heinz knew. One group of children answered a hard opaque (substitution-sensitive) question: 'does Heinz know there's a present in the box?' Another group of children answered an easier opaque question: "Does Heinz know that the ball is a present?" Thus, in addition to a transparent question, Apperly and Robinson asked two opaque questions that differed in level of difficulty.

Apperly and Robinson compared 1 false belief and 2 opacity tasks in 4 to 6-year-old children. The false belief task was an unexpected contents task, acted out with Heinz the puppet. Using a box with a picture of a teddy on the outside (which actually contains

a toy horse), children were asked a reality question: So, what's in the box? Then they were asked one false belief question: "Heinz hasn't seen inside this box before. When he first sees it, before he opens it what will he think is inside?"

There were two opacity tasks and a control task acted out with a puppet using the mental verb "know": 1) Duck/Toy Story: Dual identity object referent with two visible identities (duck /toy); this was a control or screening task and no score was given; 2) Dice/Eraser Story (Misleading Appearance Task): Dual identity object referent with one visible identity (dice) and one invisible identity (eraser); and 3) Ball/Present Story: Dual identity object referent with one visible identity (ball) and one invisible identity (present). One example is given below and detailed descriptions of the other two tasks are found in Appendix B:

Ball/Present Story:

Children are first allowed to look inside the box and the visually-obvious identity is agreed upon (children will have no difficulty in naming the ball). Next, the second identity is demonstrated. In the case of the ball, the children are conspiratorially informed that '...this is going to be a present for Heinz, except we haven't told him and we don't want him to find out right now, so we'll have to whisper very quietly when he looks...' Children hear the referential use of both labels and the mode of perceptual access (visual) is made clear. Next the children observe as Heinz looks inside the box. The lack of other perceptual access is emphasized in the case of the ball by whispering the questions.

Questions:

With the box still open, the children are asked: Q1 (transparent/extensional) can Heinz see the present inside the box? (Correct = yes). The other half has no question, guarding against the possibility of this initial 'extensional' question promoting later errors.

After the box is closed the children are asked either: Q2 (easy opacity) does Heinz know the ball is a present? (Correct = no) or Q3 (hard opacity) does Heinz know there's a (present) in the box? (Correct = no).

The four tasks were partially counterbalanced: the unexpected contents task always came first or last and the control opacity task always came before the ball and dice tasks. The order of the ball and dice tasks was counterbalanced and the question type (easy or hard) varied between participants. That is, for the ball and dice tasks for each child the order was: easy opacity question only, transparent and easy opacity questions, hard opacity only, and transparent and hard opacity questions. Results were based on the number of children in each age group (4 and 5 years) who got 0, 1 or 2 correct responses on two types of questions; the easy and hard opaque questions. Thirty-three adult participants also served as controls.

Apperly and Robinson found that all children (4 and 5 year olds) performed perfectly on the duck/toy control task. On the false belief task involving the deceptive box task (i.e., Teddy bear box with plastic horse inside) 38% of 4- year- olds and 76% of 5- year- olds answered correctly. On the opacity tasks (dice/eraser and ball/present) all children answered correctly to the transparent question (Can Heinz see the eraser/present inside the box? correct = yes). The 4- year-olds performed poorly on both question 2 (i.e., Does Heinz know the dice is an eraser/ball is a present?) and question 3 (i.e., Does Heinz

know there's an eraser/present in the box?). The 5-year-olds showed a significant difference in performance between questions 2 and 3 with question 2 being easier. In comparison to the false belief deceptive box task, question 3 but not question 2 was significantly harder than the false belief task. Comparisons across age showed that neither question 2 nor question 3 improved significantly with age.

In summary, Apperly and Robinson (1998) found a strong dissociation between children's performance on false belief and opacity tasks; the performance on false belief task improved with age, whereas performance on the hard opacity question remained poor across the two age groups. Because they failed to find a relation between false belief and opacity tasks, they concluded that their results were not compatible with an abrupt conceptual change account that proposes false belief tasks measure children's representational ability; only the opacity tasks measure representational understanding. In a second experiment they ruled out the possibility that order effects might have contributed to these results.

Apperly and Robinson (1998) proposed that children's difficulties on opacity tasks are related to their inability to distinguish between the other person's representation based on partial knowledge and their own representation based on full knowledge of a situation. Additionally, these researchers suggest that children's success on false belief tasks requires a basic ability to analyze mental representations whereas referential opacity tasks require a more sophisticated ability to represent that mental representations can be partial if the referent is only known (to the other person) under one of two or more names/descriptions. On this view, it is the representation of the other person's partial representation (knowledge) that makes opacity tasks harder than false belief tasks.

Furthermore, these researchers maintain that false belief tasks do not measure metarepresentation. They maintain that if false belief tasks measured metarepresentation, then children who pass false belief tasks should also be able to pass referential opacity tasks (Apperly & Robinson, 2002). This conclusion does not consider the possibility that both tasks are metarepresentational in nature, but opacity is more difficult because it deals with partial representations (Kamawar & Homer, 2000).

Kamawar and Olson (1999) proposed that opacity and false belief tasks are conceptually related and both require an understanding of representation and metarepresentation. They compared performance on 3 false belief and 3 opacity tasks in 3-to-7-year-old children. The standard false-belief tasks were:

- 1) An unexpected contents task using a Crayola box that contained string. There was a question for 'self' and a question for 'other': the children were asked what they had thought was in the box before it was opened and what a friend, who hadn't seen inside the box, would think was in it (not always in this order). The correct answer for both questions was 'crayons'. The child got one point for correctly remembering his/her own false belief and another point for correctly predicting his friend's false belief.

- 2) A change of location task in which a story character hid her toy turtle in a red box, only to have her brother move it to a yellow box after she went out to play. Children were then asked "Where will Linda look for her toy turtle?" A correct response would be that the girl will look for the turtle where she placed it, in the red box.

- 3) A change of location task, in which a puppet (Leo the Lion) hid a star under a white block and then went to his room to sleep. While he was sleeping, a different puppet

(Ross the monkey) moved the star and placed it under a black block. Children were asked, “Where will Leo look for his star?” A correct answer was that Leo would look under the white block. Children’s scores were summed across the false belief tasks to get a total false belief score out of a possible 3.

As a control, to ensure that errors on false belief can be attributed to lack of false belief understanding and not failure to recall previous answers, one true belief task was administered:

Children are presented with a closed bandaids box and asked what is inside:

Look at this... What's in here? [Answer: bandaids]

Let's open it and have a look. What are they? [Correct: bandaids]

Some brief conversation about bandaids...

Then... Let's play a trick on X (friend). Let's take the bandaids out and put pennies in here. [Do so and close box]

(When box is closed) What's in the box? [Correct: pennies]

True belief memory question:

What did you think was inside the box BEFORE we opened it? [Correct: bandaids]

Opacity Tasks:

Kamawar and Olson used three stories to measure performance on transparent and opaque contexts. Children were shown a model of a situation and told a story that was acted out using Playmobil® figurines and toys. Next, they were asked comprehension questions and referential opacity questions. An example of the kinds of stories used by Kamawar and Olson is given below:

Police Officer: On the street there was a Police Car. By the car, the Police Officer was tying his shoes.

Then he got up and started walking to a red house. Mark was walking home from school and he walked down that street. He saw the Police Officer drop a set of keys on the sidewalk. The Police Officer did not notice that he dropped the keys.

When Mark got to where the keys were, he picked them up and said, "I should give these keys to the Police Officer".

The Police Officer walked back to where he had come from. Mark walked over to the Police Officer. Mark said, "Hi, you dropped your keys". Mark couldn't tell that the Police Officer was Sue's dad. Mark gave the Police Officer the keys, and the Police Officer said, "Thank you".

Comprehension Question:

Did Mark give the keys to Sue's dad? [Yes]

Referential Opacity Question:

Does Mark know that he gave the keys to Sue's dad? [No]

The first question acted as a control to see if children could understand that substitution of co-referential terms is allowed in transparent contexts. The purpose of the second question was to determine whether the participants understood that Mark did not know that he gave the keys to Sue's dad because he did not know that the police officer was Sue's dad (opacity question). Thus, the second question tested whether children understood that, in referentially opaque contexts, substitution was

not allowed. Children were given a score of 1 for each story if they were able to both allow substitution in transparent contexts and not allow it in opaque contexts. Their scores across the three stories were summed to give a total opacity score; a total score out of a possible 3. Kamawar and Olson reasoned that this total score was a good measure of children's ability to deal with opaque contexts because it reflects a participant's ability to deal differently with transparent and opaque contexts. Hence, it is children's total score (based on both transparent and opaque contexts) that addresses the issue of being able to handle opacity competently and knowing when substitution is/isn't allowed. The order was counterbalanced for each set of three tasks, the two sessions and the questions within tasks. No order effects were found.

As expected, Kamawar and Olson found that performance on all tasks (false belief, transparent, and opaque contexts) improved with age. Specifically, false belief tasks improved with age and a sharp improvement was noted at 4 years of age. Performance on the transparency questions also improved with age, but not as much as for false belief tasks. Using the total opacity scores, the researchers noted that overall if children were correct on one item it was the transparent question. Total opacity scores were lower than false belief scores. The total opacity score improved with age with the most notable improvement at 5 years of age and no difference in performance between children aged 5 and 7 years. On the transparent question, 3-year-olds were able to refer to the same object in many ways, but on the opacity question they had a lot of difficulty understanding that knowing an object under one description did not mean knowing it under an alternative description. By 5 years of age children are able to understand that co-referential terms can't be used in opaque

contexts. Importantly, false belief tasks and opacity tasks were significantly correlated even after age was partialled out.

Kamawar and Olson (1999) concluded that knowledge of an object under one description does not necessarily ensure knowledge of it under all descriptions. Each description has a specific representation that must be considered separately from the referent itself. Furthermore, false belief tasks require the ability to differentiate two different objects or locations, whereas referential opacity tasks require differentiation between different descriptions of the same object, which is more difficult.

Because these researchers found a relation between performance on false belief and opacity tasks even after age was partialled out, they propose that the two types of tasks are not merely independent tasks with different levels of complexity. They argued that if both false belief and opacity tasks measured only children's ability to think about thoughts the two tasks would be indistinguishable, but they are not. In false belief tasks children think about thoughts. However, in opaque contexts they must not only think about thoughts, the different names/descriptions of objects must also become objects of thought. Opacity tasks are more complex than false belief tasks and hence more difficult. But, some common factor(s) are present besides differences in complexity. Kamawar and Olson suggest that one common factor is metarepresentational ability. On this view, although false belief tasks require a more basic and less sophisticated metarepresentational ability than opacity tasks, the two tasks are measuring the same underlying competence, namely, metarepresentational ability.

The findings of the three groups of researchers can be summarized as follows. Both deVilliers and Fitneva (1996) and Kamawar and Olson (1999) found that false

belief and opacity tasks are related. But the latter researchers found that this relation holds even after effects due to age are partialled out whereas in the deVilliers and Fitneva study, this relation was not significant once age was partialled out. DeVilliers and Fitneva (1996) found evidence that demonstrates some sort of equivalence in performance on false belief and opacity tasks. Furthermore, they argue that mastery of complementation drives the development of false belief understanding. In contrast, Kamawar and Olson (1999) found evidence of better performance on false belief than on opacity tasks. Unlike deVilliers and Fitneva, they do not argue that mastery of complementation drives false belief development or vice versa. Instead, Kamawar and Olson (1999) argue that false belief tasks measure children's ability to represent other people's representations and opacity tasks measure this ability plus the ability to separately represent different names/descriptions of the same referent. That is, both tasks measure the same underlying ability: the ability to metarepresent.

Recall that Apperly and Robinson (1998) did not find a relation between false belief and opacity tasks and concluded that false belief tasks do not measure metarepresentational ability. Kamawar and Olson (1999) disagree that false belief tasks are not metarepresentational because they did find a significant relation between the two categories of tasks. The latter researchers suggest that Apperly and Robinson did not find a relation because they only administered one false belief task and two opacity tasks. Hence, Kamawar and Olson (1999) argued that the Apperly and Robinson study was not sensitive enough to detect a relation between the tasks.

Close inspection of the two categories of tasks may help determine why Kamawar and Olson (1999) found that opacity tasks are harder than false belief tasks for the

younger children. In general, false belief tasks require children to distinguish truth from falsity. In contrast, opacity tasks require children to understand that different descriptions may represent a referent in different ways. Thus, an understanding of the representational nature of language is required; that is, children require some sort of metalinguistic ability (Kamawar & Homer, 2000). In general, metalinguistic awareness is the ability to consider the structure of language itself (syntax) as well as understanding the meanings of words (semantics) in the particular contexts. That is, the child must consider the way the object is referred to (linguistic name/description) and therefore represented, separately from the referent itself.

Kamawar (2000) compares the similarities and differences between the two categories of tasks: False belief and referential opacity tasks are similar in that the child participant has full knowledge of the situation or object and is required to answer questions about what a story character does or knows. In a false belief task the story character has incorrect knowledge about a situation (e.g., the character doesn't know chocolate was moved from the blue to the green cupboard) whereas, in an opacity task the story character has partial knowledge about an object (e.g., character knows the object referent is a ball but doesn't know it is also a present). Thus, to succeed on the false belief task children must distinguish between an accurate and inaccurate representation of the situation (two different locations for the same object) and predict how another person will behave based on the truth or falsity of that person's belief. In contrast, on the opacity task they must distinguish between accurate but partial representations (two different names or descriptions of the same object) and predict how another person will behave based on

that person's accurate, partial knowledge. Dealing with partial knowledge, they argue, is more difficult.

Kamawar (2000) also notes that some, but not all, false belief tasks use mental verbs such as 'know' or 'think'. In contrast, referential opacity tasks always use a mental verb. By definition a mental verb denotes opacity by expressing a propositional attitude that can change the truth value of the sentence. In our example, "Does Heinz know there's a present in the box?", the child must understand that Heinz has a mental attitude (i.e., that of 'knowing') towards the referent when represented as a 'ball' but not when represented as a 'present'. Thus, some metalinguistic awareness is required. In opacity tasks, therefore, children must understand the implications of mental verbs whereas, this same understanding is not necessary in false belief tasks (e.g., Maxi's chocolate story) that do not contain mental verbs. This comparison highlights the similarities and differences between the two categories of tasks (false belief and opacity).

In summary, three different groups of researchers present different views and arguments to explain children's performance on false belief tasks and referential opacity tasks: 1) the ability to handle the syntactic structure of complementation links performance on false belief and referential opacity tasks, with children who pass false belief tasks passing opacity tasks and children who fail false belief tasks failing opacity tasks (deVilliers & Fitneva, 1996; deVilliers & Pyers, 2002); 2) there is a strong dissociation between performance on false belief and opacity tasks and false belief tasks do not measure metarepresentational ability, with children who pass false belief tasks also passing referential opacity tasks (Apperly & Robinson, 1998); and 3) children perform better on false belief than on opacity tasks and false belief tasks measure a basic

metarepresentational ability, and whereas opacity tasks measure a more sophisticated metarepresentational ability, both tasks measure metarepresentational ability (Kamawar & Olson, 1999; Kamawar, 2000). Clearly, these are divergent findings for the nature of the relation between performance on false belief and opacity tasks.

Present Research

The most important issue arising from these studies is the lack of consensus among the researchers about the relation between false belief and referential opacity tasks with regard to children's performance on the two categories of tasks. Factors that may have contributed to the diverse results from the different studies are sample size, children's ages, and power. DeVilliers and Fitneva (1996) had 31 children aged 3-6 years participate in their study, Apperly and Robinson (1998) had 69 children aged 4-5 years and Kamawar and Olson (1999) had 120 children aged 3-7 years. Neither group of researchers referred specifically to power in their studies. However, of the three studies the Kamawar and Olson study had the largest number of children with the greatest age range.

The different researchers all used variations of the extensively studied standard false belief tasks. Hence, their false belief tasks are comparable. In contrast, the opacity tasks have some differences. The DF (DeVilliers and Fitneva) and KO (Kamawar and Olson) tasks use illustrations whereas the AR (Apperly and Robinson) tasks use a puppet and containers (different boxes) to tell the stories. Although two types of critical questions (transparent and opaque) are asked in each of the three types of opacity tasks, questions are placed at different points in the stories for the different tasks. These superficial differences may have contributed somewhat to the diverse findings. However,

hundreds of false belief tasks with many variations in stimuli did not affect successful performance on those tasks. Thus, task differences other than apparently superficial ones may have influenced performance on the opacity tasks.

Among these limited number of studies conducted on referential opacity in children, different kinds of referents and verbs have been used. For example, people and objects have been used as referents with different descriptors: in a misleading appearance task, an object that looks like a die is also an eraser (one visible and one invisible description, Apperly & Robinson, 1998); a ball that is big and yellow (a dual identity under two visible descriptions, Kamawar & Olson, 1999); a ball that is also a birthday present (a dual identity under one visible and one invisible description, Apperly & Robinson, 1998); an opaque container that has one visible identity (silver box) based on appearance, one visible identity (candy) based on the contents and one invisible identity (birthday present) based on the function (deVilliers & Fitneva, 1996); and people with social and role descriptors (e.g., a police officer who is also Sue's dad; Kamawar & Olson, 1999). Although these different versions of opacity tasks are broadly similar in theme, they vary in complexity and level of difficulty. For example, unlike the KO and AR referential opacity stories described earlier, the DF stories appear more complex because there are three identities (silver box/candy/birthday present); one refers to the contents, one to the container, and the third to the function. Furthermore, only one or two examples of a particular type of task were used in some experiments (e.g., only one false belief task in Apperly & Robinson, 1998). These methodological and task differences and differences in power may have contributed to the mixed results and diverse

explanations put forth by the researchers in the study of opaque contexts (Kamawar, 2000).

Hence, the primary purpose of the present research was to compare carefully chosen tasks from experiments conducted by each of these researchers to examine performance on false belief and opacity tasks within the same group of children. False belief tasks were compared to opacity tasks and the different opacity tasks were compared to each other. This type of comparison may help to integrate the findings and provide a coherent picture of how the false belief and opacity tasks are related.

Recall that the false belief tasks chosen for the present experiment are similar because they are modeled after standard false belief tasks. One false belief task was selected from each of the researchers' experiments for a total of three. Three trials of referential opacity tasks from each of the three researchers' experiments were administered for a total of nine trials on opacity tasks. With the exception of DF opacity tasks, all of the referential opacity tasks used dual identity object referents with one visible and one invisible name/description. The DF opacity tasks had one visible and two invisible names/descriptions. Although the action verbs varied (e.g. put, find, and throw) only the mental verb 'know' was used. This consistency is important because using different kinds of verbs and referents changes task difficulty. For example, action referent tasks are easier than object referent tasks (Kamawar, 2000). Similarly, tasks using the mental verb 'know' are easier than tasks using the mental verb 'think' (Kamawar, 2000). Tasks will be compared for level of difficulty and complexity.

Jenkins and Astington (1996) found that general language ability and verbal memory were significant predictors of false belief understanding after effects due to age were

partialled out, but nonverbal memory was not. Furthermore, they suggest that either the general language measure (test of early language development or TELD) or the verbal memory measure (sentence memory measure of the Stanford-Binet) is “sufficient to take account of the effects of language on false belief understanding” (Jenkins & Astington, 1996, p.5). However, the authors acknowledge that a finer grained measure of memory may account for more variance than the general language measure.

A number of cognitive factors are important in young children’s performance on false belief tasks. For example, deVilliers and Pyers (2002) administered a battery of tests to examine the relation between language and understanding of false beliefs. They found that the best predictor of success on false belief tasks was not general language ability (as suggested by Astington & Jenkins, 1999). Rather, they found that mastery of tensed complements, a specific aspect of syntax, was the best predictor of false-belief understanding. Thus, a complementation comprehension task developed by deVilliers was administered in the present study.

In addition to a short term memory task, a working memory task may provide information about the type of skill needed to succeed on false belief tasks. Working memory refers to a mechanism that allows us to understand and mentally represent the immediate environment and to retain information about immediate past experience in order to accomplish a cognitive task (Baddeley & Logie, 1999). Tests of working memory include span tasks that require both temporary storage and manipulation of information. For children to succeed on false belief tasks, they have to hold in memory all of the necessary information (their own and other’s representations) while simultaneously comparing differences in the representations and predicting the other

person's behaviour. Success on false belief tasks appears to involve more than the measure of short term memory used on the Stanford Binet in the recall of increasingly complex sentences (used by Jenkins & Astington, 1996). Therefore, in the present research, children were administered both a short term memory (forward digit span) and a working memory (backward digit span) task to examine the relation between performance on false belief and opacity tasks. The forward digit span task measures immediate recall ability. The backward digit span task measures working memory ability (Gathercole & Pickering, 2000). The purpose of the two memory tasks is to account for any contribution of memory to performance on false belief and opacity tasks. Measures of working memory span are highly correlated with measures of language comprehension, vocabulary learning, reasoning, reading comprehension and general achievement and ability (Engle, Kane & Tuholski, 1999). Thus, to reduce the total number of tasks administered, other cognitive tasks (e.g., TELD) were not considered here. Although there are many span tasks available, the forward and backward digit span tasks are used in this experiment because they were used previously by Kamawar and Olson (2000) and they are standardized tests (see Appendix D). However, norms have not been established for children under the age of seven years.

Recall that different researchers suggest that four abilities are necessary for children to succeed on referential opacity tasks: (1) metarepresentational ability as demonstrated by passing false belief tasks (i.e., ability to represent another person's knowledge and beliefs); (2) the ability to handle the syntactic structure of complementation (DeVilliers & Fitneva, 1996; deVilliers & Pyers, 2002) as measured by the complementation task; (3) the ability to understand that another person's representation of an object or situation may

be restricted to one description in the absence of knowledge about alternative descriptions (i.e., partial knowledge) as measured by opacity tasks (Apperly & Robinson, 1998; Kamawar & Olson, 1999, Kamawar, 2000).

The primary goal of the present research is to replicate previous findings and extend the number and variety of false belief and opacity tasks to be compared across the same group of children. Examining performance of the same children across a range of opacity and false belief tasks has not been done previously and provides a powerful design to determine the strength and complexity of relations among the tasks. In addition, the design has the potential to highlight differences across opacity tasks that may have led to the diverse results reported by the different researchers. Importantly, the present research design allows for extension and integration of the existing research on children's developing sensitivity to referentially opaque contexts.

Many children around 4 years of age who demonstrate metarepresentational ability by passing false belief tasks fail the referential opacity tasks (Apperly & Robinson, 1998; deVilliers & Fitneva, 1996, Kamawar & Olson 1999, 2000). Yet, by 5 years of age most children are able to pass false belief tasks and by age 7 most children pass opacity tasks. Thus, for the present research children between 3; 9 and 6; 11 years of age were recruited.

Hypotheses

1) Consistent with previous research, it was expected that children's performance would improve with age on both the false belief and referential opacity tasks (de Villiers & Fitneva, 1996; Kamawar & Olson, 1999; Kamawar, 2000).

2) Kamawar and Olson (1999) used the largest number of false belief and opacity tasks for comparison in the largest sample of children. Therefore, based on the study with the largest sample, I predict that for the younger children, performance on false belief tasks will be better than on referential opacity tasks as previously shown (Kamawar & Olson, 1999).

3) Performance on false belief tasks will correlate significantly with referential opacity tasks as previously shown (Kamawar & Olson, 1999; Kamawar, 2000).

4) The three different types of opacity tasks have not been compared previously in the same group of children. Because Kamawar and Olson (1999) found that false belief and opacity tasks measured the same underlying metarepresentational ability, I predict that performance across the three different types of opacity tasks (DF, AR, KO) will show different levels of difficulty but will be correlated even when age is controlled.

Method

Participants

Permission was sought at local daycares and schools (junior and senior Kindergarten, and grade one classes) to recruit 60 English-speaking children. Across the age range, children were divided into three age groups with 20 in each age group. There were 11 girls and 9 boys in the youngest group (3;9 to 4;9 years), $M = 4; 4$ years, $SD = 3.53$. The middle group had 9 girls and 11 boys (4;10 to 5;10 years), $M = 5; 2$ years, $SD = 3.30$. There were 10 girls and 10 boys in the oldest group (5;11 to 6;11 years) $M = 6;6$ years, $SD = 2.68$. See Appendix E for the information letter to parents and the informed consent form.

Experiment Design

Sixty children received FB, DF, and KO tasks. All of them received two types of questions (1 transparent and 1 opaque) at the DF level and KO level of opacity (20 children in each of the 3 age groups for a total of 60). However, at the AR level of opacity, although children also received transparent and opaque questions, 30 received an easy opaque question and 30 received a hard opaque question. They all received the same transparent question. Thus, thirty children received the transparent question and an easy opaque question (Group A). There were 10 children in each of the 3 age groups for a total of 30 children. Another thirty children received the same transparent question and a hard opaque question (Group B). Again, there were 10 children in each of the 3 age groups for a total of 30 children. In sum, all children received the same false belief, digit span, complementation comprehension, KO opacity and DF opacity tasks (within subject). However, in addition to these tasks, half of the participants also received only the AR A opacity tasks while the other half received only the AR B opacity tasks (AR A/AR B was between subjects). See Appendix F for the experimental design layout.

Materials

The 16 tasks are included in full in the appendices. One true belief screening task and three false belief tasks (Appendix A) were administered. Opacity tasks (Appendix B) included 3 trials (i.e., stories) from each of three types (Apperly & Robinson, deVilliers & Fitneva, and Kamawar & Olson) for a total of nine trials. One language task (complementation comprehension) was administered (Appendix C). Two memory tasks (forward and backward digit span) were administered to assess short term and working memory respectively (Appendix D). In an effort to be concise the task names are abbreviated in some places: TB = true belief, Comp = complementation, FB = false

belief; Apperly and Robinson = AR; deVilliers and Fitneva = DF; Kamawar and Olson = KO.

True Belief Control Task

One true belief task was administered first as a screening task. The purpose of this task was to ensure that children could recall their own earlier beliefs. Thus, any errors made on false belief questions could be attributed to lack of understanding false belief and not due to either a failure to recall earlier utterances or to a bias to answer with reality. The true belief task was given at the beginning of session one and no score was given for this task. If a child failed this task, then their session one data was discarded and they did not continue participating in the study. Children who passed this task continued participating in the remaining two sessions.

In the true belief task, children were presented with a closed box of band-aids and asked what is inside. Then, to play a trick on someone, the band-aids are removed and replaced with pennies. The children are asked: “What’s in the box?” [Correct: pennies], and a true belief memory question, “What did you think was inside the box BEFORE we opened it?” [Correct: band-aids].

False Belief Tasks

Three false belief tasks were administered to measure children's ability to recognize false beliefs in themselves and in others. Children earned a maximum of 5 points for the false belief tasks (2 for self and 3 for other questions). Two were unexpected contents tasks and one was a change of location task (see Appendix A for complete stories).

- (1) An unexpected contents task using a box with a picture of a teddy on the outside which actually contained a toy horse (a question for self and a question for other = 2 points) (Apperly & Robinson, 1998).
- (2) An unexpected contents task using a Crayola box filled with stickers (a question for self and a question for other = 2 points) (Kamawar, 2000).
- (3) A change of location task in which a story character puts a cake in a cupboard, and while he is out playing, his Mom moves the cake to the refrigerator (question for other = 1 point) (deVilliers & Fitneva, 1996).

Table 1

Summary of scores for false belief tasks

FB Task	Point Value
Change of Location DF	1
Unexpected Contents AR	2
Unexpected Contents KO	2
Range of possible scores	0, 1, 2, 3, 4, 5 (max = 5)

Referential Opacity Tasks

Nine referential opacity trials (3 trials for each type of opacity task) were administered. The mental verb “know” was used for all of the referential opacity questions. On each trial, one point was given for each story if the participant correctly answered both the transparent and opacity questions thus demonstrating that substitution of a co-referential term was allowed in transparent questions but not in opacity questions. A score was computed by summing each child’s score on each set of 3 trials (per type): DF (deVilliers and Fitneva) score, AR (Apperly and Robinson) score, and KO (Kamawar

and Olson) score. These scores were then used to compare across opacity tasks, compute correlations across opacity tasks, and compare opacity to false belief performance. One trial from each of the three different researchers' experiments are described below (the full set of trials can be found in Appendix B).

DeVilliers and Fitneva Tasks

On each of three different trials, children were told three different stories with coloured illustrations. The child participant and one of the story characters had full knowledge of the situation (Mom) whereas the other story character (little girl) had only partial knowledge. The container was opaque, therefore the object referent had one visible identity (silver box) based on appearance, one visible identity (candy) based on the contents and one invisible identity (birthday present) based on the function. An example of a transparent question is: "Did the mom put the silver box on the top shelf?" (Correct = yes). Two opacity questions were asked: (e.g. "Does the little girl know the silver box (standard referent) is on the top shelf?" Correct = yes, and "Does the little girl know the candy/birthday present (co-referential terms) is on the top shelf?" Correct = no).

Apperly and Robinson Tasks

On each of three different trials, children were told three different stories that were acted out with a puppet. In each story, dual identity object referents are used with one visible (e.g. ball) and one invisible identity (e.g. present). An example of a transparent question is: "Can Stephen see the present inside the box?" (Correct answer = yes). Two types of opacity questions were asked. In version "A" the referential opacity question was asked (e.g. "Does Stephen know the ball is a present?" Correct answer =

no). In version “B” the referential opacity question was asked (e.g. “Does Stephen know there is a present inside the box?” Correct answer = no).

Kamawar and Olson Tasks

On each of three different trials, children were told three different stories with coloured illustrations. In each story, dual identity object referents were used with one visible (e.g., candy) and one invisible identity (e.g., treat). An example of a transparent question is: “Did Mark find a candy?” (Correct answer = yes). An example of referential opacity questions are: “Does Mark know that he found a candy?” (Correct answer = yes). “Does Mark know that he found Beth’s treat?” (Correct answer = no).

Table 2

Summary of scores for trials on opacity tasks

Opacity Scores
DF: up to a total of 3
AR: up to a total of 3
KO: up to a total of 3
Range of scores: 0, 1, 2, 3

Complementation Comprehension Task

In this task children were shown 8 sets of photographs with 2 photographs in each set. While pointing to the photographs one at a time, the experimenter read a statement about each photograph. Then the child answered a question about what a character “told” someone in the picture (what- tell question) or what a character thought someone was thinking about in the picture (what-think). For example, in the what- tell situation the researcher read, “She told her husband she saw a ghost, but it was really a blanket.” The

child's task was to answer the question "What did she tell her husband?" In the what-think situation the researcher read "She thought the girl was reading a book, but she was really playing cards." Then the child was asked "What did she think?" There were 4 what-think questions and 4 what – tell questions for a total of 8 points. To be counted as correct, children were only required to answer with the appropriate content and not necessarily produce a complement structure as this was a comprehension task. The communication verb structure (the what-tell score out of 4) will be used in the analysis as the complement measure for reasons described earlier.

Memory Tasks

The forward and backward digit span sub-tests from the Stanford Binet intelligence tests were used to test the child's short term and working memory abilities (See Appendix D). These sub-tests were chosen because they are standardized tests and have been used previously in research on opaque contexts (Kamawar, 2000). In the forward digit span test, children were asked to remember a series of single digit numbers and repeat them back (e.g., experimenter says "5, 8" and the child says "5, 8"). In the backward digit span test, children were asked to remember a series of single digit numbers and say them backwards (e.g., experimenter says "5, 8" and the child says "8, 5"). Sequence lengths were increased and the task was administered until the child made errors on 3 or more out of four on two consecutive pairs of items. Frequently researchers use the combined raw scores on the forward and backward digit span in their analyses. However, as noted in Gathercole and Pickering (2000), this practice has been criticized because there is evidence to suggest that forward digit span measures memory for short term recall only whereas backward digit span measures the ability to simultaneously store

and manipulate information. Therefore, the individual scores for forward and backward digit span were analyzed separately in the present experiment. Memory measures were partialled out of other correlations because the purpose of the present experiment is to show that the relations between FB and opacity tasks and among different types of opacity tasks are determined by children's metarepresentational ability and not just their ability to remember the information.

Procedure

Each child was interviewed individually in three 15 – 20 minute sessions over a one to two week span. The female researcher introduced herself to the children as a university student who was interested in how children learn. They were told that she would read stories and ask them questions about what was going on in the stories to help the researcher with her homework. The children were also told they could go back to class if they did not want to hear any stories. After each session (or if they decided not to continue) the child was given a choice of some stickers and thanked for their participation. The child was then returned to the classroom.

Over three sessions the children received the following 16 tasks: one true belief screening task, three false belief tasks (1 DF, 1 AR, 1 KO), nine trials on opacity tasks (3 DF, 3 AR, 3 KO), one language task (complementation), and two memory tasks (1 short term and 1 working memory). At the beginning of session one, the true belief screening task was administered. In each session, the child received one false belief and three opacity tasks (1 DF, 1 AR, 1KO). In session one the language task was given and in session two the forward digit span task was given. In the final session, the working memory task was given in between two opacity tasks or between an opacity and false

belief task. The working memory task was expected not to be as enjoyable as the other tasks for some children, thus it was given in the last session and followed by a more enjoyable task (i.e., an opacity or false belief task).

Tasks were counterbalanced with two fixed orders of presentation and order of the questions was counterbalanced for DF tasks. Counterbalancing for question order was not needed for KO tasks because previous research had ruled out order effects in these tasks (Kamawar & Olson, 1999, 2000). For AR tasks, counterbalancing for question order was not required because these tasks were a between subjects design: the easy opacity question was given to 30 of the 60 children and the hard opacity question was given to the other 30 children.

Table 3

Summary of tasks in order one and order two

Task Order 1	Session 1	Session 2	Session 3
	True Belief	KO candy	AR ball
	DF silver box	FB/Crayons	FB/Cake
	FB/Teddy	DF black bag	KO truck
	AR dice	Digit Span (F)	Digit Span (B)
	Comp	AR car	DF red tin
	KO allowance		
Task Order 2	Session 1	Session 2	Session 3
	True Belief	KO candy	DF silver box
	AR ball	FB/Crayons	FB/Teddy
	FB/Cake	DF black bag	AR dice
	KO truck	Digit Span (F)	Digit Span (B)
	Comp	AR car	KO allowance
	DF red tin		

Sessions were tape recorded to ensure accuracy regarding scoring of the data.

Results

Participants were recruited from daycare centers and elementary schools in middle class neighbourhoods around the Ottawa area. Fifteen children were excluded for

a variety of reasons: 6 children failed the true belief task, 3 did not want to participate because they were too shy, 5 did not complete the three sessions (4 due to absence, one child was not keen possibly because English was his second language) and 1 preferred to play. Sixty children (30 boys and 30 girls) between 3; 9 and 6; 11 years of age participated in the study and received 16 tasks over three sessions for a total of 960 stories /tasks. The tasks were administered in two orders with half of the children receiving order 1 and the other half receiving order 2 (see Table 3, p. 43).

Recall that in Condition A, 30 children received Apperly and Robinson's transparent question and their easy opaque question (AR A). In Condition B, 30 children received Apperly and Robinson's transparent question and their difficult opaque question (AR B). Unless otherwise stated, all of the results are reported in terms of Condition A and Condition B of Apperly and Robinson tasks. Thus, Condition A refers to those participants who receive false belief, digit span, complement comprehension, KO, DF and AR A, and Condition B refers to those participants who receive false belief, digit span, complement comprehension, KO, DF, and AR B. First, the results are reported for replication of the findings of each of the three groups of researchers. Then results are reported across all the tasks.

Types of Analyses

The main statistical tests used were ANOVA, partial correlations, Chi-square test for independence, and multiple regressions. Children's performance was analyzed using their raw summed scores for the ANOVAs, partial correlations, and multiple regressions. Recall that for false belief tasks children received a summed score out of 5 (summed across the three false belief tasks) and children received a summed score out of 3 for each

opacity type (ARA, AR B, KO, and DF). In order to determine whether passing opacity tasks required false belief understanding, (i.e., whether passing opacity tasks is contingent upon passing false belief tasks) Chi-square analyses were performed. Summed scores were converted to categorical data (pass/fail). On false belief tasks a pass was a score of 4 or 5 out of 5 and for opacity tasks a pass was a score of 2 or 3 out of 3. Results were the same for a more stringent test (pass = 5/5 for false belief and 3/3 for opacity tasks) unless otherwise stated.

Replication of deVilliers and Fitneva Tasks

In 1996, deVilliers and Fitneva found that performance on false belief tasks did not correlate significantly with their opacity tasks once age was partialled out. This result was replicated in the present study; controlling for age, performance on false belief tasks did not correlate with DF opacity tasks, $r(57) = .02$, n.s. Previously these researchers argued for some kind of equivalence relation between false belief and their opacity tasks even though they failed to find a significant correlation between the two kinds of tasks once age was partialled out: children who passed false belief tasks also passed DF opacity tasks and children who failed false belief tasks failed their opacity tasks. However, contrary to deVilliers and Fitneva's previous findings, in the present study a comparison of performance on DF opacity tasks and false belief tasks revealed that performance on false belief tasks is independent of performance on DF opacity tasks, $\chi^2(1, n = 60) = 1.66$, n.s. (see Table 4).

Table 4.

Performance on false belief and DF opacity tasks

		False Belief		Total
		pass	fail	
DF	pass	28	8	36
	fail	15	9	24
	Total	43	17	N = 60

Examination of question order was applicable only to the deVilliers and Fitneva opacity tasks. A one-way ANOVA (using summed scores for the DF opacity tasks as the dependent variable and DF version as the factor denoting question order) revealed that contrary to deVilliers and Fitneva's findings, there was a significant main effect of order, $F(1, 59) = 4.18, p < .05$. Performance on DF question order version A was significantly better than on DF question order version B ($\bar{M} = 1.93$ and $\bar{M} = 1.37$, $SD = 1.11$ and $SD = 1.03$ respectively). Implications of this finding are addressed in the discussion.

In summary, consistent with deVilliers and Fitneva's findings, the DF opacity tasks were not correlated with false belief tasks (controlling for age). However, evidence for an equivalence relation between DF tasks and false belief tasks was not found in the present study. Taken together, these two findings suggest that false belief tasks are not related to DF opacity tasks and passing false belief tasks is not a prerequisite for passing DF opacity tasks as previously claimed.

Replication of Apperly and Robinson Tasks

The main screening task used for this experiment was the true belief task. It was the first task administered and children who failed this task finished session one but did not participate in sessions two or three and their session one data were destroyed. A second control task (duck/toy) used by Apperly and Robinson was also administered in the present study. In their original study, Apperly and Robinson reasoned that if children could not pass the simpler duck/toy control task they could not have any real understanding of the opacity task. In other words, they argued that children are probably just guessing if they get the opacity task right and the simpler duck/toy control task wrong. In the present study, 17 children who passed the true belief screening task did not pass the Apperly and Robinson duck/toy control task. Most of these children were in the young and middle age groups but there were also a few in the oldest age group. It is interesting to note that 29% (5 out of 17) of the children who failed the Apperly and Robinson control task obtained a perfect maximum score on the AR opacity stories. Compared to the original Apperly and Robinson study, the present study contained a higher criterion for passing an opacity task (children had to get both the transparent and the opacity questions correct to get a score of 1 for an opacity story). Furthermore, children received three opacity stories compared to two in the original Apperly and Robinson study. It seems unlikely that these children were guessing given the higher criterion for passing an opacity task and the addition of another opacity task in the present study. Passing the duck/toy control task may indicate that children have the basic understanding necessary to engage in opacity tasks, but failing it does not seem to indicate that they do not.

To avoid losing such a large number of participants (17) it was decided to keep children in the present study who passed the true belief task but failed the Apperly and Robinson control task (especially given that a number of children who failed the AR control went on to get 3/3 on the AR opacity questions). However, for analyses aimed at replicating the Apperly and Robinson results, the 17 children who failed their control tasks were excluded. Fortunately, the excluded cases were evenly distributed across the two AR versions (AR A and AR B). Thus, the number of children included in the present analyses is 43 compared to 69 in Apperly and Robinson's 1998 study.

In the present experiment, performance on the AR false belief task was consistent with Apperly and Robinson's previous results. Thirty three percent of the children in the youngest group (aged 3; 9 to 4; 9), 54% of the children in the middle group (aged 4; 10 to 5; 10), and 89% of the children in the oldest group (aged 5; 11 to 6; 11) answered correctly to both a question for self and a question for other on the unexpected contents (teddy bear/horse) false belief task (see Table 5). In the present study this was a more stringent test of false belief understanding than in the original Apperly and Robinson study because in addition to a question for "self" ("What did you think was inside the box before you opened it?"), children were asked a question for "other" ("What would a friend, who had not seen inside the box before, think was inside the box before (s) he opened it?").

Apperly and Robinson examined performance on the two opacity questions within age groups using Chi-square analyses to determine whether or not there was a significant difference in difficulty between the AR A (easy) opacity question and the AR B (hard) opacity question. Recall that in the present study, the Apperly and Robinson tasks are

also a between subjects design. Children are asked the AR A (easy) opaque question in Condition A whereas they are asked the AR B (difficult) opaque question in Condition B. In the present study, Fisher's Exact Test was used because the use of Chi-square resulted in some cells with expected counts that were very small. This test is a test for independence in a 2 X 2 table. On false belief tasks a pass was a score of 4 or 5 out of 5 and for opacity tasks a pass was a score of 2 or 3 out of 3.

Consistent with Apperly and Robinson's previous results, the youngest children did poorly on both opacity questions. Indeed, none of the 8 children in the youngest group passed the easy AR A opacity questions and none of the 4 children in the youngest group passed the easy AR B hard opacity questions. Also, consistent with their previous results, more of the oldest children who had the AR A question passed the opacity task than those who had the AR B question, $\chi^2(1, n=18) = 5.84$, Fisher's Exact Test = .05. However, all of the children did poorly on the hard AR B question; there was a floor effect (see Table 5). Nonetheless, these results are reported so that the current study can be directly compared to Apperly and Robinson's (1998) results.

Table 5
Numbers (%) of children who answered questions correctly on AR tasks

	False Belief	AR A	AR B
Young (3; 9 – 4; 9 yrs.)	4/12 (33%)	0/8 (0%)	0/4 (0%)
Middle (4; 10 – 5; 10 yrs.)	7/13 (54%)	4/8 (50%)	1/5 (20%)
Old (5; 11 – 6; 11 yrs.)	16/18 (89%)	7/9 (78%)	1/9 (11%)

Apperly and Robinson used the McNemar test to compare their false belief task to their two opacity questions. This test is a nonparametric test for two related dichotomous

variables and tests for changes in responses using the chi-square distribution. Typically, a significance value less than 0.05 is considered significant. To compare results of the present study to the Apperly and Robinson study, the same test was used.

Consistent with their findings, comparison of the false belief task to the hard AR B opaque question showed a significant difference in difficulty, McNemar's $\chi^2 (1, n = 18) = .53, p = .001$. Of the 12 children who passed the false belief task, only 1 child passed the difficult AR B question. Clearly, the false belief task is easier than the AR B opacity question which produced the floor effect described above.

However, contrary to their findings, comparison of the false belief task to the easy AR A opaque question also showed a significant difference in difficulty, McNemar's $\chi^2 (1, n = 18) = 3.72, p = .001$, Fisher's Exact Test $p = .129$ (recall, this test is a test for independence in a 2 X 2 table).

Table 6

Performance on false belief and AR A opacity tasks

		False Belief		Total
		pass	fail	
AR A	pass	11	0	11
	fail	8	6	14
	Total	19	6	N = 25

None of the 6 children who failed the false belief tasks passed the AR A opacity question. Of the 19 children who passed the false belief task, 58% passed the AR A question and 42% failed the AR A question. Hence, passing false belief tasks may be

necessary but is not sufficient for passing AR A opacity tasks. Previously, Apperly and Robinson argued that their opacity tasks were not related to false belief tasks. However, in the present study a relation was found between the AR A opacity tasks and false belief tasks thereby contradicting their findings.

Comparisons across age groups were also done for both the AR A and the AR B questions. Contrary to Apperly and Robinson's findings, performance on the AR A easy question improved significantly with age, $\chi^2(1, n = 25) = 8.18, p < .02$. However, consistent with their findings, performance on AR B hard question did not improve with age, $\chi^2(1, n = 25) = .44, n.s.$ Clearly, the hard AR B question was too difficult for all of the children and the results show a floor effect. Therefore, analyses involving the AR B question generally demonstrate null findings and will not be mentioned except when they are directly related to a replication or an earlier stated hypothesis/research question.

In summary, consistent with Apperly and Robinson's (1998) findings children's performance on the AR false belief task improved significantly with age and children found false belief tasks easier than both the opacity questions (AR A and AR B). Contrary to their previous findings, children's performance on the AR A easy opacity question was related to false belief tasks and children's performance improved significantly with age on this question. Consistent with their previous findings, children's performance on the AR B difficult opacity question was not related to their false belief task and did not improve significantly with age.

Replication of Kamawar and Olson Tasks

Performance on false belief tasks was hypothesized to correlate significantly with referential opacity tasks as previously shown (Kamawar & Olson, 1999; Kamawar,

2000). Thus, one-tailed partial correlations were performed. Controlling for age, forward and backward digit span, false belief tasks correlated significantly with KO opacity tasks $r(55) = .26, p < .03$, one-tailed. This result is consistent with Kamawar and Olson's previous findings.

Children's performance on false belief and KO opacity tasks was examined by comparing the number of children who failed and passed in each type of task; $\chi^2(1, n = 60) = 10.59, p < .001$ (see Table 7). Thus the two tasks are related. Of the 17 children who failed false belief tasks, only two passed the KO opacity tasks. Of the 43 children who passed the false belief task, 58% passed the KO opacity tasks and 42% failed the KO tasks. Hence, passing false belief tasks may be necessary but is not sufficient for passing KO opacity tasks. This result is consistent with Kamawar and Olson's previous findings.

Table 7

Performance on false belief and KO opacity tasks

		False Belief		Total
		pass	fail	
KO	pass	25	2	27
	fail	18	15	33
	Total	43	17	N = 60

Consistent with previous findings, for the KO tasks, performance varied significantly with age groups, $F(2, 57) = 23.35, p < .001$. Tukey HSD post hoc tests showed that on KO tasks the youngest (3; 9 to 4; 9 years) children's performance differed significantly from both the middle (4; 10 to 5; 10) group $p < .001$, and the oldest (5; 11 to

6; 11) group, $p < .001$. However, the middle group did not differ significantly from the oldest group.

In the second hypothesis, the younger children were predicted to perform better on false belief tasks than on KO referential opacity tasks. To enable the comparison of false belief to opacity scores, false belief scores were scaled by multiplying them by .6 so that their maximum score would be the same as for opacity scores (i.e., out of three). A repeated measures ANOVA revealed a main effect of task (false belief vs. KO), $F(1, 57) = 61.87, p < .001$, ($M = 2.35, SD = .94$, and $M = 1.30, SD = 1.18$, for false belief and KO opacity respectively). Also, there was a main effect of age group, $F(1, 57) = 31.87, p < .001$. There were no significant interactions. To see if the results of the ANOVA held for each age group, paired samples t-tests (with the file split by age group) were computed. Recall that a 'pass' for false belief was a 4 or better out of 5 (now scaled to be out of 3), and a 'pass' for opacity was 2 or better out of 3. This result revealed that false belief scores were significantly higher than KO opacity scores for each of the three age groups. Thus as predicted in the second hypothesis, the younger children performed better on false belief tasks than on KO referential opacity tasks. Furthermore, this finding held for the other two age groups (see Table 8).

Table 8

Mean performances and standard deviations on false belief vs. KO opacity tasks per age group (maximum score = 3) N = 60

	<u>False Belief</u>		<u>KO Opacity</u>		t
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Young	1.59	1.11	0.20	0.41	5.73***
Middle	2.55	0.67	1.70	1.08	3.77***
Old	2.91	0.29	2.00	1.03	4.04***

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

This analysis was also done for each of the other opacity types (DF, AR A and AR B). The means and standard deviations comparing false belief to each of the opacity types across age groups are summarized and presented in Table 9.

In summary, KO opacity tasks were significantly correlated with false belief tasks (controlling for age), and passing false belief tasks was necessary but not sufficient for passing KO tasks. Children's performance on KO opacity tasks improved significantly across age groups and performance on false belief tasks was significantly better than on KO opacity tasks. Hence, all of Kamawar and Olson's previous findings were replicated in the present experiment.

Comparison of False Belief to Opacity Tasks

In the second hypothesis the younger children were predicted to perform better on false belief tasks than on referential opacity tasks. This prediction was supported earlier for KO tasks. To enable the comparison of false belief to opacity tasks, false belief scores

were scaled by multiplying them by .6 so that their maximum score would be the same as for the opacity scores (i.e., out of 3). To determine if this result would also hold for the DF tasks, the same analyses were done. A repeated measures ANOVA revealed a main effect of task (false belief vs. DF), $F(1, 57) = 19.19, p < .001$. Also, there was a main effect of age group, $F(1, 57) = 27.65, p < .001$. There were no interactions. To see if results of the ANOVA held up for each age group, paired samples t-test (with the file split by age group) were computed. They revealed that false belief scores were significantly higher than DF opacity scores for each of the three age groups. Thus as predicted in the second hypothesis, the younger children performed better on false belief tasks than on DF referential opacity tasks. Furthermore, this finding held for the other two age groups (see Table 9).

Similarly, to determine if this result would also hold for the AR A tasks the same analyses were done. A repeated measures ANOVA revealed a main effect of task (false belief vs. AR A), $F(1, 27) = 15.86, p < .001$. Also, there was a main effect of age group, $F(1, 27) = 11.28, p < .001$. There were no interactions. To see if results of the ANOVA held up for each age group, paired samples t-test (with the file split by age group) were computed. They revealed that false belief scores were significantly higher than AR A opacity scores for each of the three age groups. As discussed in the AR replication section, performance on false belief tasks was significantly better than on the AR B tasks because none of the children in either age group could do the AR B tasks.

Thus, as predicted in the second hypothesis, the younger children, performed better on false belief tasks than on all of the referential opacity tasks DF, AR A, AR B, and KO.

Table 9

Mean performances and standard deviations on false belief vs. opacity tasks per age group (maximum score = 3) N = 60 for KO and DF; N = 25 for AR A)

	False Belief		KO Opacity		t
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Young	1.59	1.11	0.20	0.41	5.73***
Middle	2.55	0.67	1.70	1.08	3.77***
Old	2.91	0.29	2.00	1.03	4.04***

	False Belief		DF Opacity		t
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Young	1.59	1.11	0.85	1.04	2.15*
Middle	2.55	0.67	1.85	0.93	2.65*
Old	2.91	0.29	2.25	0.85	3.25**

	False Belief		AR A Opacity		t
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Young	1.50	1.14	0.60	0.97	1.71
Middle	2.58	0.64	1.50	1.27	3.02*
Old	2.94	0.19	2.10	1.10	2.70*

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

Figure 1. Mean scores for false belief tasks and each type of opacity task (DF, AR A, AR B, and KO) across age groups.

Mean Differences across Opacity Tasks

It is important to note that the three different types of opacity tasks have not been compared previously in the same group of children. Each of the studies that have examined children's performance on opaque contexts seems to have found different levels of difficulty (i.e. Apperly & Robinson, 1998; deVilliers & Fitneva, 1996; Kamawar & Olson, 1999). Thus, the fourth hypothesis states that differences in difficulty across the types of opacity tasks would be expected in the present study. In order to see which of the opacity tasks differed from one another, planned paired samples t-tests were performed. They showed significant mean differences only between DF and AR B opacity tasks, $t(29) = -3.97, p < .001$, two-tailed, and for DF vs. KO opacity tasks, $t(59) = 2.09, p < .04$, two-tailed. These significant mean differences on opacity scores show that these tasks differ in levels of difficulty (AR B: $M = .87, SD = .82$ vs. DF: $M = 1.73, SD = 1.05$) and

(DF: $\underline{M} = 1.65$, $SD = 1.10$ vs. KO: $\underline{M} = 1.30$, $SD = 1.18$). There were no significant mean differences for AR B vs. KO, or for AR A vs. KO or for AR A vs. DF (see Figure 1).

Hence, differences in difficulty were found between the easiest (DF) and two most difficult (AR B and KO) opacity tasks. That is, children found the DF tasks easier than the AR B and KO tasks. These results are partly consistent with the fourth hypothesis; different levels of difficulty were present but only for certain tasks.

Age Effects

Performance on false belief tasks across age groups.

False belief scores were analyzed in a 2 (task order: 1 vs. 2) x 2 (gender: female vs. male) x 3 (age group: young vs. middle vs. old) between subjects ANOVA. There were no significant main effects of order and gender and there were no interactions. However, as hypothesized, there was a main effect of age group, $F(2, 48) = 14.78$, $p < .001$.

A one way ANOVA was done to determine whether performance improved with age. As expected and consistent with previous research, performance on false belief tasks improved significantly with age, $F(2, 57) = 15.88$, $p < .001$. Significant results from post hoc tests are summarized in Table 10. The youngest (3; 9 – 4; 9 yrs.) children's performance on false belief tasks differed significantly from both the middle (4; 10 – 5; 10 yrs.) group and the oldest group (5; 11 – 6; 11 yrs.). However the middle group did not differ significantly from the oldest group.

Age effects and opacity tasks.

As hypothesized, there was a main effect of age on opacity tasks for both Condition A and Condition B ($F(2, 19) = 19.77$, $p < .001$, and $F(2, 19) = 5.61$, $p < .01$,

respectively). In order to determine whether this result held up for each of the opacity types, separate one-way ANOVA and Tukey HSD post hoc tests were done to examine effects of age for each. Age effects for KO tasks were reported earlier in the ‘Kamawar and Olson’ section and are included in Table 10 which provides a summary of age effects for false belief and the three types of opacity tasks.

Performance on AR A tasks improved significantly with age, $F(2, 27) = 4.55, p < .02$. Tukey HSD post hoc tests showed that the youngest (3; 9 to 4; 9 years) children’s performance on AR A tasks differed significantly only from the oldest (5; 11 to 6; 11) group $p < .015$. Although Apperly and Robinson did not do this particular analysis (they used only Chi-square analyses), this result does not support their finding of no improvement across age groups for their easy (AR A) question.

For the DF tasks, performance also differed significantly across age groups, $F(2, 57) = 11.62, p < .001$. The youngest (3; 9 to 4; 9 years) children’s performance on DF tasks differed significantly from both the middle (4; 10 to 5; 10) AR A and the oldest (5; 11 to 6; 11) group. However, the middle group did not differ significantly from the oldest group.

In summary, age effects were similar across the age groups for all of the tasks; there were significant differences between the young vs. middle group (except for AR A) and young vs. old group. There were no significant differences in performance between the middle and old groups for any of the tasks. See Table 10 for a summary of significant age effects across false belief and opacity tasks.

Table 10
Significant age effects across false belief and opacity tasks

	<u>Young vs. Middle</u>		<u>Young vs. Old</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
FB	2.65 vs. 4.25	1.84 vs. 1.12	2.65 vs. 4.85	1.84 vs. 0.49
AR A			0.60 vs. 2.10	0.97 vs. 1.10
DF	0.85 vs. 1.90	1.04 vs. 0.97	0.85 vs. 2.25	1.04 vs. 0.85
KO	0.20 vs. 1.70	0.41 vs. 1.08	0.20 vs. 2.00	0.41 vs. 1.03

These results are significant at $p < .01$. $n = 20$ in each age group except for AR A which had $n = 10$ per age group.

Comparisons of Previous to Present Findings

Recall, in order to determine whether passing opacity tasks required false belief understanding, (i.e., whether passing opacity tasks is contingent upon passing false belief tasks) Chi-square analyses were performed and reported above for each of the three different types of opacity tasks. A summary of past and present findings is presented in Table 11.

Table 11
Previous and Present Findings for False Belief and Opacity Tasks

	<u>Previous Research</u>	<u>Present Findings</u>
FB and DF	equivalent FB = DF	not related FB easier than DF
FB and AR A	not related FB easier than AR A	related FB easier than AR A
FB and AR B	not related FB easier than AR B	not related FB easier than AR B
FB and KO	related FB easier than KO	related FB easier than KO

Comparisons Across Opacity Tasks

In order to determine whether or not opacity performance was similar across the types, opacity scores (out of 3) were analyzed, in a 2 (task order: 1 vs. 2) x 2 gender (female vs. male) x 3 (age group: young vs. middle vs. old) x 3 (opacity type: AR vs. DF vs. KO) ANOVA with repeated measures on the last factor. For Condition A (children received AR A, DF, KO, digit span, complement comprehension and false belief tasks), there were no significant main effects of opacity, order, and gender. However, there was a three-way interaction of opacity x order x gender, $F(2, 19) = 3.22, p < .05$. For Condition B (children received AR B, DF, KO, digit span, complement comprehension and false belief tasks), there were no significant main effects of order and gender. However, there was a main effect of opacity, $F(2, 19) = 4.74, p < .01$. These main

effects are qualified by a three-way interaction of opacity x age group x order, $F(2, 19) = 2.74, p < .04$.

Three-way Interaction of Opacity x Order x Gender

Recall, tasks were counterbalanced with two fixed orders of presentation (see Table 3 p.43). The pattern of performance across opacity tasks (AR A, DF, and KO) for females was similar for both orders but better on order one than on order two. The pattern of performance across these same tasks for males was different from the females (see Figure 2 below).

Figure 2. Mean opacity scores as a function of type of opacity task, gender, and order.

For the males, the pattern of performance on order two was similar to the female performance on both order one and order two. Performance for the boys in order one stood out as being different from the boys' performance on order two and the girls' performance on both orders. Inspection of the mean ages, mean scores on false belief and mean scores on backward digit span was done in an attempt to shed some light on the nature of the interaction. The means for age, false belief scores, and backward digit span were similar across order and gender (see Table 12). Thus, there is no obvious explanation for the interaction from examining the relevant variables expected to make a difference in performance. However, this may be explained in part because of the unequal numbers in the cells across the two orders. For order one, there were 9 females and 9 males whereas for order two the composition was 10 boys and 2 females.

Table 12
Mean Performances and Standard Deviations on Age, False Belief, Digit Span B for Gender and Order

	<u>Order 1 Boys</u>		<u>Order 2 Boys</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Age	63.56	10.35	61.80	12.66
False Belief	3.67	1.94	3.60	1.65
Digit Span (B)	2.22 (n = 9)	1.09	1.60 (n = 10)	1.27
	<u>Order 1 Girls</u>		<u>Order 2 Girls</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Age	69.00	10.32	73.5	10.61
False Belief	4.22	1.39	5.00	0.001
Digit Span (B)	2.56 (n = 9)	1.24	2.50 (n = 2)	0.71

Three-way Interaction of Opacity x Age Group x Order

Close examination of this interaction shows that for both orders mean performance across age groups was similar in the young and middle groups of children. A different pattern of performance emerged for the older group of children across the two orders (see Figure 3).

Figure 3. Mean opacity scores as a function of type of opacity task, order, and age group.

Mean performance for DF tasks was almost the same for order one as for order two. In contrast, for both the AR B and KO tasks mean performance was better for order one than for order two. For AR B tasks, mean performance declined from order one to order two. For KO tasks, this change in performance was even more dramatic, from order one to order two (see Table 13).

Table 13
Mean Performances and Standard Deviations on Opacity Tasks for Order 1 and Order 2

	<u>Order 1</u>		<u>Order 2</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
AR B	1.30	1.21	1.03	1.21
DF	1.63	1.10	1.67	1.10
KO	1.57	1.22	1.03	1.22

This pattern of performance across orders may be explained by differences in task difficulty and the benefits of practice. As reported above, paired samples t tests showed significant mean differences in difficulty between AR B and DF tasks, and DF vs. KO tasks. Furthermore, order one began with the easiest of the opacity tasks (DF) and children had experience (practice) with this task before encountering the more difficult AR B and KO opacity tasks. Order two began with the difficult AR B task and unlike order one, children were unable to benefit from their experience with this task. For the easiest (DF) tasks order did not affect performance as much as on the difficult tasks. For the difficult tasks, prior experience with an easier similar task enhanced performance on the more difficult tasks for order one. This combination of order and task difficulty differentially affected performance across the different age groups.

The pattern of decline in performance across the three age groups was different for order one compared to order two. It was the oldest age group that contributed most to this decline from order one to order two for the KO tasks ($\underline{M} = 2.60$, $SD = .52$ in order

one to $\bar{M} = 1.40$, $SD = 1.07$ in order two). The means and standard deviations are summarized and presented in Table 14.

Table 14
Means and Standard Deviations for Opacity by Age and Order

	Order 1		Order 2	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
AR B young	0.90	1.20	0.50	0.53
AR B middle	1.30	1.16	1.00	1.05
AR B old	1.70	1.25	1.60	1.07
DF young	0.70	0.95	1.00	1.15
DF middle	2.10	0.74	1.60	1.07
DF old	2.10	0.99	2.40	0.70
KO young	0.20	0.42	0.20	0.42
KO middle	1.90	0.99	1.50	1.18
KO old	2.60	0.52	1.40	1.07

Practice Effects

Analyses of practice effects were conducted by comparing paired samples of mean opacity scores (out of 3) for session 1 vs. session 3 within order 1 and session 1 vs. session 3 within order 2. Within order 1, there were no significant practice effects for either Condition A (KO, DF and AR A tasks) or Condition B (KO, DF, and AR B tasks) $t(17) = -.356$, n.s. and $t(11) = .29$, n.s. respectively. Within order two, a different pattern

of results emerged. For Condition A, there were no significant practice effects $t(11) = -1.33$, n.s. However, for Condition B there were significant practice effects, paired t-tests showed $t(17) = -3.19$, $p < .005$, two-tailed. Performance on session three was significantly better than performance on session one, $\underline{M} = 1.72$, $SD = 1.02$ and $\underline{M} = .83$, $SD = .79$, respectively. Close inspection of the age groups revealed that for both orders, age groups were evenly distributed across the two orders. That is, there were 10 children in each age group (young, middle, and old) in order one and in order two. Furthermore, the mean ages in each order were similar: young ($\underline{M} = 50.75$ vs. $\underline{M} = 51.50$ mos. for orders 1 and 2, respectively), middle ($\underline{M} = 64.25$ vs. $\underline{M} = 62$ mos. for orders 1 and 2, respectively), and old ($\underline{M} = 78$ vs. $\underline{M} = 77$ mos. for orders 1 and 2, respectively). Hence, practice effects for order 2 in Condition B were not due to having more children in the older age group in one of the orders. Task difficulty may have contributed to the practice effects. The AR B task was the most difficult of the opacity tasks and showed a floor effect. In order two, the AR B opacity task is encountered first whereas in order one it is the third task introduced. Starting with a difficult task may have negatively affected performance on the remaining tasks.

Intercorrelations among the Tasks

In order to determine the degree to which children's performance on various tasks correlate with one another, Pearson correlations were computed. Given that performance on all of the tasks (except AR B) correlates highly with age, partial correlations controlling for age (computed in months) were calculated. Results are reported for tasks that all children received ($N = 60$, see Table 15). When computing partial correlations (controlling for age in months) for AR A and AR B tasks, the 17 children who did not

pass their Apperly and Robinson control task were excluded because Apperly and Robinson would not have included them in their study. Results for AR A tasks ($N = 25$) and AR B tasks ($N = 18$) are reported in Table 16.

Relations between false belief and opacity tasks.

Partial correlations (controlling for age) were computed to determine whether performance on the opacity tasks was related to performance on false belief tasks. Importantly, the KO opacity tasks correlated significantly with false belief tasks (as reported earlier in the replication section). However, performance on DF, AR A, and AR B opacity tasks did not correlate significantly with false belief tasks. The non-significant results for partial correlations between AR tasks and false belief tasks should be interpreted with caution due to the small number of participants included in the analysis of both AR A and AR B tasks. Contrary to the third hypothesis, performance on false belief tasks did not correlate significantly with referential opacity tasks overall. Only the KO tasks were related to false belief tasks.

Relations among opacity tasks.

The fourth hypothesis states that differences in difficulty across the different types of opacity tasks would be expected in the present study (some were found as reported earlier). However, it was also hypothesized that even if the opacity tasks were ordered by complexity, there would nonetheless be an underlying similarity in performance across them that would be revealed by inter-correlations. Thus partial correlations (controlling for age) were computed. Performance on DF and KO opacity tasks ($N = 60$) did not correlate significantly. However, AR A tasks correlated significantly with KO opacity tasks ($N = 25$, see table 16). Given the small number of participants in the AR A analysis,

this latter result seems robust. The AR B tasks did not correlate significantly with any of the other tasks because of the floor effect. Thus, not only was performance poor on the AR B task, it is not related to the other tasks. Contrary to the fourth hypothesis, performance across the three different types of opacity tasks (DF, AR A, AR B, and KO) was not consistently correlated when age was controlled. Only KO and AR A opacity tasks were significantly correlated.

Table 15
Partial correlation matrix: age partialled out

	False Belief	DF	KO	Digit Span (B)	Digit Span (F)	Comp Tell	Comp Think
1. False Belief	--						
2. DF	.02	--					
3. KO	.26*	.06	--				
4. Digit Span (B)	-.09	-.04	-.005	--			
5. Digit Span (F)	.07	.25	.09	.19	--		
6. Comp Tell	.53***	-.07	.04	.16	.11	--	
7. Comp Think	.58***	-.03	-.08	.20	.07	.94***	--

Note: All correlations have N = 60. * $p < .05$ ** $p < .01$ *** $p < .001$

Table 16
Partial correlation matrix: age partialled out for AR A and AR B

	False Belief	DF	KO	Digit Span (B)	Digit Span (F)	Comp Tell	Comp Think
1. AR A	.19	-.15	.72***	.06	.02	-.01	-.004
2. AR B	-.17	.07	-.31	-.43	.01	.16	.08

Note: AR A correlations have N = 25 and AR B correlations have N = 18.

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

Complementation task.

Recall that children were also administered a language measure: complementation comprehension for ‘think’ and ‘tell’. Also recall that deVilliers (1996) argued that the ability to deal with syntax of complementation was related to false belief and opacity understanding. To determine if the deVilliers & Pyers (2002) finding was replicated, two-tailed partial correlations (controlling for age) were computed. Results showed that performance on the two types of complementation questions (“what think” and “what tell”) were highly correlated to one another and correlated significantly with false belief tasks, thus replicating deVilliers and Pyers results. However, complementation performance was not related to any of the opacity tasks (i.e., DF, KO, AR A, and AR B).

Memory tasks.

Controlling for age, performance on the memory tasks (forward and backward digit span) was not significantly correlated with false belief or opacity tasks. Thus, performance on the false belief and opacity tasks was not influenced by memory constraints.

Contributions of Unique Variance to each type of Opacity Task

To determine whether performance on false belief tasks and the complementation –tell task contributed a significant amount of unique variance to each of the three types of opacity tasks, hierarchical regressions were performed for opacity tasks that had significant raw correlations (i.e., correlations with nothing partialled out) with the false belief and the complementation task. Significant raw correlations are summarized and presented in Table 17.

Table 17
Raw correlations among all the tasks.

	Age (months)	AR A	AR B	DF	KO	False Belief	Comp -tell	Digit Span (B)	Digit Span (F)
Age (months)	--								
AR A	.52**	--							
AR B	.14	a	--						
DF	.49**	.19	.20	--					
KO	.64**	.67**		.36**	--				
False Belief	.59**	.38*	.18	.30*	.56**	--			
Comp-tell	.46**	.35	.28	.17	.32*	.66**	--		
Digit Span (Backward)	.60**	.40*	-.02	.27*	.38**	.29*	.39**	--	
Digit Span (Forward)	.56**	.31	.22	.45**	.42**	.38**	.39**	.46**	--

'a' - there was no variability for AR B thus no value was computed

Recall that deVilliers & Fitneva maintained that for children to understand other people's mental states, it is necessary for them to understand the language structure that refers to mental states. Children's performance on the "tell" questions examines their ability to handle a specific aspect of complex syntax (tensed complements), without confounding this ability with use of mental verbs (see p.12). Therefore, the complementation task using the mental verb 'think' (i.e., "what-think" questions) was not included in the regressions.

Separate hierarchical regressions were performed (using summary scores out of 3 for each type of opacity task) with DF, KO, and AR A, as the dependent variables and with backward digit span (raw scores), age in months, complementation-tell (out of 4) and false belief (out of 5), as independent variables. For each regression, to determine if false belief accounts for a significant amount of unique variance, forward and backward digit span and age in months were entered first, then complementation-tell, with false belief entered last. Similarly, to determine if performance on the complementation "tell"

question accounts for a significant amount of unique variance, forward and backward digit span and age were entered first, then false belief, with the complementation “tell” performance entered last. As mentioned above, the complementation “tell” task was used so as to not confound syntactic structure with the use of mental verbs.

With KO as the dependent variable, and forward and backward digit span, age in months, complementation-tell and false belief as independent variables, significant changes were noted for age and for false belief. Thus, false belief performance accounts for a significant amount of unique variance in performance on KO opacity tasks even after variability due to age and performance on forward and backward digit span, and complementation-tell were removed (see Table 18). In contrast, for both the DF and AR A opacity tasks, there were no significant R square changes other than age for the first step (see tables 19 and 20 respectively). As none of the measures correlated with AR B (floor effect), no regressions were computed.

Table 18

Summary of hierarchical regression analysis predicting KO referential opacity from age, forward digit span, backward digit span, comp-tell, and false belief (N = 60).

Model/Step/Variable	B	SE _B	β	R ² or R ² Δ
Both Models				
1. Digit Span (F)	0.07	0.11	0.09	
Digit Span (B)	0.02	0.13	- 0.02	
Age	0.06	0.02	0.60	0.41***

Model – 1				
2. False Belief	0.18	0.09	0.24	0.04*
3. Comp-tell	0.25	0.11	0.03	0.001

Model - 2				
2. Comp-tell	-0.13	0.13	-0.14	0.01
3. False Belief	0.25	0.11	0.33	0.05*

Note. Reported regression coefficients are from final step. * $p < .05$, ** $p < .001$

Table 19

Summary of hierarchical regression analysis predicting DF referential opacity from age, forward digit span, backward digit span, comp-tell, and false belief (N = 60).

Model/Step/Variable	B	SE _B	β	R ² or R ² Δ
Models-1				
1. Digit Span (F)	0.21	0.11	0.28	
Digit Span (B)	-0.09	0.13	-0.09	
Age	0.04	0.02	0.39	0.29***
2. False Belief	-0.01	0.10	-0.01	0.001

Note. Reported regression coefficients are from final step. * $p < .05$, ** $p < .001$

Table 20

Summary of hierarchical regression analysis predicting AR A referential opacity from age, backward digit span, and false belief (N = 60).

Model/Step/Variable	B	SE _B	β	R ² or R ² Δ
Model-1				
1. Digit Span (B)	0.10	0.23	0.10	
Age	0.05	0.02	0.46	0.28*
2. False Belief	0.05	0.18	0.06	0.002

Note. Reported regression coefficients are from final step. * $p < .05$, ** $p < .001$

Summary of the Main Findings

In the present study, relations between false belief and each type of opacity task were explored using partial correlations, multiple regressions and Chi-square analyses. Results from these analyses revealed that only KO opacity tasks were significantly related to false belief tasks. Also, a main effect of opacity was found for condition B (children received false belief, digit span, complement comprehension, DF, KO, and AR B tasks). Mean differences between opacity tasks were found for DF vs. KO and DF vs. AR B tasks. However, across types of opacity tasks, only performance on AR A and KO tasks was significantly correlated. Hence, except for AR A and KO tasks, the opacity tasks are fundamentally different from each other. How well each of the opacity types measures performance on opaque contexts will be explored in the discussion.

Discussion

The main purpose of the present research was to investigate relations between false belief and referential opacity tasks. The tasks used were from previous studies by three different groups of researchers: deVilliers and Fitneva (1996), Apperly and Robinson (1998), and Kamawar and Olson, 1999. Interestingly, each of these studies found different relations between false belief and opacity understanding. In addition, they also found differing levels of performance on their opacity tasks. Taken together, these findings seemed to indicate that though all of the researchers claimed to be examining children's ability to deal with opaque contexts, there was something different about the opacity tasks that they were using. In this thesis, one false belief task and 3 opacity tasks were drawn from each of the three studies to allow for direct comparisons. Hence, 3 false belief and 9 opacity tasks were given to sixty children aged 3;9 to 6;11 years.

Very few researchers have compared false belief to opacity tasks. Further, comparing false belief tasks to three similar but different types of referential opacity tasks within the same group of children has not been done before. Thus, the most unique and important contributions of the present research are the comparisons of children's performance: (1) between false belief and each of the three types of opacity tasks; and (2) across the three types of opacity tasks (i.e., to see whether the three types of opacity tasks were tapping the same underlying ability).

Relations among False Belief and the Referential Opacity Tasks

Children's performance on false belief tasks was compared to their performance on the three different types of opacity tasks (DF, AR, and KO) using partial correlations, multiple regression and Chi-square analyses. The results from these analyses are clear and consistent. Contrary to the third hypothesis, performance on false belief tasks did not correlate significantly with the referential opacity tasks overall (once age was controlled). The relation of false belief to each type of opacity task is discussed next.

False belief and KO opacity tasks.

Of the three types of opacity tasks (DF, AR, and KO), only the KO tasks correlated with false belief tasks. Also, the multiple regression analyses showed that false belief scores account for a significant amount of unique variance in performance on KO opacity tasks even after backward digit span, age, and the complementation 'tell' question were removed. Furthermore, consistent with both the regression analyses and the partial correlations, the Chi square results showed that the KO opacity tasks and false belief tasks are related. That is, performance on the KO opacity tasks required some understanding of false belief. These results provide strong evidence that only

performance on the KO opacity tasks shares some common factor with false belief tasks. False belief tasks are thought to measure metarepresentational ability. Therefore, based on this pattern of results, false belief and KO opacity tasks measure the same underlying metarepresentational ability. However, false belief tasks are easier than the KO opacity tasks. Therefore, false belief and KO opacity tasks measure different levels of metarepresentational ability.

To succeed on false belief tasks such as the change of location cake task, children must compare their own true belief (e.g., the cake is in the refrigerator) to the character's false belief (e.g. the cake is still in the cupboard). That is, children must represent two different representations; their own true representation of reality and the character's false representation that differs from reality. Then children must represent the difference between the two representations. Note that the object of interest (i.e., cake) cannot be in location 'x' (the refrigerator) and location 'y' (the cupboard) at the same time. In contrast, for children to succeed on opacity tasks, they must compare their own knowledge of co-referential terms that are both true to the character's partial knowledge of one co-referential term that is also true. Thus, children must represent two different representations that are true: their own based on full knowledge of co-referential terms and the character's based on partial knowledge of the co-referential terms. Note that the object of interest (e.g., ball) is both an 'x' (ball) and a 'y' (present). As with false beliefs, children must grasp the difference between the two representations. In the false belief tasks it is relatively easy to contrast the two representations because one is true and one is false (as the object can only be in one location). However, in the opacity tasks the contrast between the two representations is difficult because the difference between the

two true representations is subtle. In other words, it is more difficult for children to compare and contrast two true representations than it is for them to compare and contrast two representations one of which is true and the other false. In opacity tasks, the co-referential terms (i.e., ball/present) are both true and the children must grasp that because the character knows the referent under one description (i.e., co-referential term) does not mean that the character knows the referent under all relevant and true descriptions. The character's representation only represents one description of the referent. It is the representation of the character's accurate but partial knowledge that makes the opacity tasks more difficult than the false belief tasks (Kamawar & Homer, 2000). False belief tasks measure some basic level of metarepresentational ability and KO opacity tasks measure a more sophisticated level of metarepresentational ability that involves the ability to consider opaque contexts.

Kamawar (2000) argued that to succeed on false belief tasks children must be able to think about thoughts. For example in the change of location task children must represent what the character thinks about where the cake is located based on the character's knowledge, which is different from the child's knowledge. In contrast, to succeed on referential opacity tasks children must think about or represent the co-referential terms themselves (e.g., silver box, candy, and birthday present) as being separate from the referent (i.e., the object). Hence, children's performance is better on false belief tasks than on referential opacity tasks because in addition to thinking about thoughts, children must consider the language (co-referential terms) as an object of thought separately from the referent being described by the language. On both false belief and referential opacity tasks children have to think about the real situation and represent

how their own representation differs from that of the story character. However, unlike false belief tasks, knowledge about the language (co-referential terms) is more important than the referent itself for children to succeed on referential opacity tasks. Hence, in the past Kamawar (2000) argued that both false belief and opacity tasks measure metarepresentational ability but opacity tasks require a sophisticated level of metarepresentational ability that includes the consideration of language. The middle and oldest groups of children in the present study were at or near ceiling on the false belief tasks and have acquired a basic level of metarepresentational ability. However, these children are far from ceiling performance on the KO opacity tasks and are in the process of acquiring the more sophisticated metarepresentational ability that is needed to master opacity tasks. In summary, the KO results in the present study suggest that the level of metarepresentational ability measured by the false belief tasks is an elementary and early developing ability. In contrast, the additional abilities required to pass opacity tasks are later developing and not yet mastered by the children in the present study.

False belief and AR opacity tasks.

Partial correlations and multiple regression analyses in the present study showed that false belief tasks are not related to the easy AR A opacity tasks. However, similar to the KO tasks, Chi square analyses showed that passing false belief tasks seems to be a necessary but not sufficient prerequisite for passing KO and AR A opacity tasks. Finding that there is a relation between AR A and false belief tasks is new. The small number of participants in the AR A tasks may account for the failure to find a correlation between false belief and AR A tasks and failure to find a unique contribution of false belief performance to the AR A tasks. In the Chi-square analysis the Fisher's exact test was

designed to detect relations with small numbers of participants; thus, a significant Chi-square result was found. Clearly, some false belief understanding is necessary for successful performance on the AR A opacity tasks.

Apperly and Robinson (1998) did not find a relation between false belief and opacity tasks and concluded that false belief tasks do not measure metarepresentational ability. Kamawar and Olson (1999) disagreed that false belief tasks are not metarepresentational because of the nature of the false belief task (recall that they did find a significant relation between the two categories of tasks). The latter researchers suggested that Apperly and Robinson did not find a relation because they only administered one false belief task and two opacity tasks to each child. Hence, Kamawar and Olson (1999) argued that the Apperly and Robinson study was not sensitive enough to detect a relation between the tasks. In the present study, the addition of one AR opacity task (for a total of three) and two false belief tasks (also for a total of three) seems to have increased the ability to detect this new finding of a relation between the two types of tasks, thereby offering support to Kamawar and Olson's suggestion. Given the present results following the addition of tasks, the Apperly and Robinson (1998) study appears to have lacked the sensitivity to detect a relation between their opacity and false belief tasks. Similar to the KO results in the present study, the AR A opacity tasks can be said to measure the elementary metarepresentational ability measured by false belief tasks plus the more sophisticated level of metarepresentational ability that involves the ability to consider opaque contexts (i.e., the co-referential terms). Also similar to the KO opacity tasks, performance on the AR A opacity tasks revealed that the additional abilities

required to pass opacity tasks are later developing and not yet mastered by the children in the present study.

The AR B tasks were not related to false belief tasks and this is a replication of past findings. Most of the children (i.e., the middle and oldest groups) were at or near ceiling performance on the false belief tasks but none of the children could do the AR B task. Clearly, the metarepresentational abilities necessary to pass false belief tasks were mastered long before successful performance on the AR B opacity tasks is possible. Hence, it is not surprising that the two tasks were not related.

False belief and DF opacity tasks.

Consistent with deVilliers and Fitneva's 1996 findings, false belief tasks were not correlated to DF opacity tasks (once age was controlled) in the present study. Furthermore, Chi-square and multiple regression analysis showed that the DF tasks were not related to false belief tasks in the present study. Thus, contrary to their previous finding of some sort of equivalence relation between false belief and opacity tasks, the two tasks were not related in the present study. Hence, whether DF opacity tasks measure metarepresentational ability is open to question. In their 1996 study deVilliers and Fitneva (1996) did not explain their failure to find a correlation between their opacity tasks and false belief tasks. Also, in the present study, the DF tasks were not related to the other referential opacity tasks (i.e., AR A and KO). Therefore, a close examination of children's performance was conducted for the three DF opacity tasks and is discussed below in the 'Opacity Tasks' section. Some methodological problems were observed that may account for the failure to find a relation between the DF tasks and the false belief tasks, and between the DF tasks and the other opacity tasks.

Summary

In the present study partial correlations, multiple regressions, and Chi-square analyses revealed that false belief performance is related to the KO opacity tasks. Also, Chi-square analysis showed that false belief and AR A opacity tasks are related. Despite the small sample ($N = 25$ when children were excluded because they failed the AR control task), performance on ARA and KO opacity tasks showed a robust correlation. The AR B opacity tasks were too difficult for the children and they are just beginning to master the easy opacity tasks. Thus, performance on the difficult AR B opacity tasks would not be expected to correlate with the near ceiling performance on false belief tasks at this point in their handling of opacity tasks.

Complementation Task

Performance on both the complementation questions correlated significantly with false belief tasks (even with age partialled out) thus replicating deVilliers and Pyers (2002) results. In addition, the two types of complementation questions (“what think” and “what tell”) were highly correlated. These results support the researchers’ specific linguistic explanation for a link between complementation understanding and false belief performance. The complementation comprehension task provided a receptive measure of the ability to deal with the syntactic structures. DeVilliers and Pyers argued that this ability is a necessary but not sufficient requirement for passing false belief tasks. Most of the children in the present study were at ceiling performance on the complementation and false belief tasks. Hence, they have acquired the ability to deal with tensed complements and then false beliefs. However, these abilities appear to be necessary but not sufficient to pass opacity tasks and are mastered well before children acquire the ability to deal with

opaque contexts. Hence, the complementation task correlated with the false belief tasks but not the opacity tasks.

Opacity Tasks

Recall that there was no main effect of opacity for children who received Condition A (AR A, DF, KO and remaining tasks); thus, performance was consistent across tasks for this group of children. However, for the children who received Condition B (AR B, DF KO, and remaining tasks) there was a main effect of opacity; thus, performance was not consistent across tasks in this group of children. Significant mean differences across opacity tasks were found between AR B and DF opacity tasks, and between DF and KO opacity tasks. There were no significant mean differences between AR A and DF tasks or AR A and KO tasks. Hence, differences in difficulty were found between the easiest (DF) and two most difficult (AR B and KO) opacity tasks. That is, children found the DF tasks easier than the AR B and KO tasks. These significant mean differences show that these tasks differ in levels of difficulty. These results are partly consistent with the fourth hypothesis; different levels of difficulty were present but only for certain tasks.

The present study was designed to examine whether the three types of opacity tasks were tapping the same underlying abilities by administering the different types of tasks to the same groups of children. To determine whether the different opacity tasks tap the same underlying abilities, analyses were conducted to examine whether the tasks were related. Partial correlations showed that performance across opacity tasks (AR A, AR B, DF, and KO) was not related overall. Only the AR A and KO tasks were significantly correlated. Thus, Kamawar and Olson's referential opacity tasks are

significantly correlated with Apperly and Robinson's (easy) opacity question. Contrary to the fourth hypothesis, performance across the three different types of opacity tasks (DF, AR, and KO) was not correlated overall when age was controlled. These conflicting results demonstrate that the opacity tasks are fundamentally different from each other and have different levels of difficulty. While consistent with the literature, these results are much more striking and informative as they are based primarily on within-subject analyses.

In summary, the AR A and KO opacity tasks were the only tasks that were related. Furthermore they did not differ in level of difficulty. Thus, these two opacity tasks can be said to measure the same level of the same underlying metarepresentational ability that allows children to deal with opaque contexts. Not much can be said about differences in difficulty across the remaining comparisons among the opacity tasks because the tasks were not related. Contrary to my predictions, it is not possible to say that the tasks are related and that differences across the tasks are merely due to differences in difficulty. Since the tasks are not related then it is not possible to claim that they measure the same underlying abilities. Possible explanations for the differences in performance (i.e., difficulty) will now be addressed.

The AR B, DF and KO tasks are not related; yet, they all claim to measure opacity. In general, the opacity tasks were similar in that they all involved knowing and ignorant story characters, partial knowledge on the part of story characters, and co-referential terms. Recall from the introduction that an opaque context was defined as one in which substitution of a co-referential term may change the truth value of a sentence; for example, "Does Stephen know the ball is a present?" This question is said to be

opaque because the terms are co-referential (the object can be known as a ball or a present) and because substitutions (of co-referential terms) in the embedded proposition can affect the truth value of the sentence overall (e.g., Stephen may only know the object under the description ‘ball’ and not ‘present’). For all of the tasks, the criteria for passing opacity was the same; children had to be correct on both the transparent and opacity questions to score 1 point for that particular story. The interesting and important question is: why was performance by the same children so inconsistent across the different opacity tasks? To answer this question three different factors will be considered: (1) linguistic (structural) differences among the critical transparent and opaque questions; (2) task differences such as type and number of co-referential terms (e.g. invisible vs. visible descriptors); and (3) the representational ability that contributes to successful performance.

Linguistic differences across the critical questions are minimal. Indeed, the transparent and opaque questions were designed to be similar in structure across the different opacity tasks. Furthermore, only the mental verb ‘know’ was used for the opacity questions. Recall that with mental state verbs it is possible to have a proposition embedded under that verb. For example, the following three questions are similar in structure and the proposition is underlined, “Does Amy know the silver box is on the top shelf?” (DF), “Does Stephen know the ball is a present?”(AR A) and “Does Mark know that he found a candy?” (KO). All of these examples involve tensed complements and, therefore, the differences among them are not due to differing syntactic requirements (more examples of critical transparent and opacity questions are found in Appendix G).

Given that linguistic differences among critical questions did not likely contribute to the diverse performance across opacity tasks, other task differences must be considered.

DF opacity tasks.

The three DF opacity tasks stand out as different from the AR and KO tasks because 3 terms are used to describe the referent instead of 2. At the beginning of each DF task the children are explicitly told about the descriptors. For example, in the DF silver box story, the children are told “This silver box is a birthday present. This birthday present is candy.” The object is referred to as a silver box and a birthday present and candy. The standard referent is the silver box and it is visible in the illustrations. The other co-referential terms include candy (an invisible descriptor) and birthday present (an invisible descriptor). The term ‘birthday present’ is co-referential to the silver box and the candy; thus the term ‘birthday present’ is doubly co-referential (i.e., it co-refers to two different objects). The term ‘candy’ is a different object from the silver box and as such is not co-referential to the silver box (though it was treated as being co-referential by DeVilliers, 1996). However, the term candy is co-referential to the term birthday present. Thus, the (invisible) descriptor ‘birthday present’ is co-referential with the (visible) ‘silver box’ descriptor and with the (invisible) ‘candy’ descriptor but the term candy does not co-refer to the silver box. Hence, the DF tasks have some methodological problems. How these problems may have affected children’s performance on the DF tasks will be discussed next.

The opacity questions were asked in two sets of two questions in each of the DF opacity tasks. For example, in version A of the silver box story the first set of opacity questions (c and d) are:

(c) “Does Amy know the silver box is on the top shelf?”

(d) “Does Amy know the candy is on the top shelf?”

In contrast, for version B of the silver box story the first set of opacity questions are:

(c) “Does Amy know the silver box is on the top shelf?”

(d) “Does Amy know the birthday present is on the top shelf?”

The descriptors are confounded such that in DF version A children are only asked ‘opacity’ questions using the term ‘candy’. However, contrary to DeVilliers (1996), I argue that the question “Does Amy know the candy is on the top shelf” is not opaque because candy is not co-referential to the silver box; it doesn’t pick out the same referent with a different description, instead it picks out a different referent (the contents of the silver box). However, in version B when children are asked opacity questions using the term ‘birthday present’, the context is opaque because ‘birthday present’ is co-referential with the ‘silver box’. Thus, it turns out that Versions A and B vary more than just in question order: a careful review of the opaque contexts that DeVilliers used calls into question the opacity of one of them. This helps to interpret other results as well: as reported in the results section, children’s performance was significantly better for version A than on version B. It is reasonable that children did better on Version A given that the contexts are not truly opaque, and that this difference makes them easier.

Using different objects, the silver box story was used as a model to create two more stories that were identical in structure. Therefore, in each story, there were descriptions that referred to the container, the contents and the function of the item. In addition, the questions that were asked all followed a similar structure (see above). The opacity questions consisted of two types: (1) those using the standard description

(container) and those using a non-standard description (contents/function). The pattern of errors was consistent by question type; children's performance was the poorest on the non-standard description question when the function descriptor was used. Hence, the DF opacity tasks were easier than the AR A and KO opacity tasks because of the methodological problems described above.

Please note that additional analyses were carried out using only the version with the valid opaque context (version B). Partial correlations (controlling for age) revealed that performance on DF version B was not correlated with any of the following tasks: FB tasks, KO opacity tasks, forward and backward digit span and the complementation task. Partial correlations (controlling for age) were also computed for DF version B with the AR opacity tasks. However, with the file split by version for the AR tasks (recall this was a between subjects task with $N = 43$) there were too few children to consider the results interpretable. Hence, in the present study the overall results did not change when considering only the DF version with the valid opaque context. For future studies, the task could be modified to overcome the confound described above by eliminating the 'contents' descriptor and keeping the two terms that are truly co-referential, the container (e.g. silver box) and the function descriptor (e.g. birthday present).

KO and AR A opacity tasks.

Interestingly, the KO tasks are related to the AR A tasks and the two types of tasks are at the same level of difficulty (i.e., no significant mean differences). Comparable to the one AR A (easy opacity) question "Does Stephen know the ball is a present?" are the two KO opacity questions "Does Mark know that he found a candy?" and "Does Mark know that he found Beth's treat?". The two KO questions have the same

result as the one AR A question; the two questions simply draw the children's attention to both co-referential terms and measure their awareness of the two terms. Thus, these two types of tasks may essentially measure the same thing; children's metarepresentational ability and their ability to represent the co-referential terms separately from the referent.

AR B opacity tasks.

The difficult AR B opacity question was designed to demonstrate that important developments in children's understanding of the mind take place beyond the age of 4 years; contrary to Perner's claim, children's understanding of the mind is not an all-or-nothing event that ends at age 4. In this difficult AR task, children are shown a ball inside a box and are told it is going to be a present for the puppet Stephen....but Stephen doesn't know it is going to be a present for him... then the puppet Stephen 'looks' inside the box and the children are asked the transparent question "Can Stephen see the present inside the box?". Next the box is closed and children are asked the difficult AR B opacity question "Does Stephen know there's a present in the box?" The easy AR A task is identical to the AR B task except the opacity question is changed to "Does Stephen know the ball is a present?" The latter opacity question is easier because it draws attention to both co-referential terms thereby supporting children's representations by making it explicit that there are two different descriptions of the same referent. By having the two descriptions presented within one question, it may be easier for children to attend to the fact that a distinction may need to be made between the two (i.e., only one is a known descriptor to the character). The difficult AR B question "Does Stephen know there is a present in the box?" requires children to apply the knowledge that they have about their representation of the situation and their knowledge of the invisible co-referential term. In

other words, the question itself does not support the children by making it explicit that there are two different descriptions of the same referent. The AR B question measures both the elementary metarepresentational ability and the ability to consider co-referential terms as objects of thought separately from the referent. However, unlike the AR A tasks, the children are required to make the connection on their own that the two different descriptions refer to the same object (referent); the AR B question does not make the difference between the two descriptions explicit. This raises important questions for consideration in future research.

Directions for Future Research

Thus far, false belief tasks are thought to measure early developing metarepresentational ability; the KO and AR A opacity tasks (i.e., easy opacity questions) measure a somewhat later developing metarepresentational ability that includes the ability to consider the co-referential terms as objects of thought. However, children in the present study did better on the AR A and KO than on the AR B tasks because the opacity question(s) supported the formation of their representations in the former but not the latter tasks. Comparison of the difficult AR B opacity question across opacity tasks (e.g. AR B, KO, and DF) may be a better measure, than the easier opacity questions of when children have mastered the ability to handle opaque contexts. The important question is, at what age will children master opacity questions (i.e., like those in the AR B tasks) that do not explicitly support the formation of their representations?

In a future research project, the answer to this question may be found by comparing children's performance on the AR B opacity question to performance on modified versions of the KO and DF tasks. Essentially, the AR B task involves a

transparent question and an opacity question, both of which use the invisible descriptor to refer to the object (e.g. ball/present): transparent question “Can Stephen see the present (invisible descriptor) inside the box?” and opacity question “Does Stephen know there’s a present inside the box?” Specifically, the KO opacity tasks could be modified by using only the transparent and opacity questions that use the invisible descriptor (e.g. transparent question “Did Mark find Beth’s treat?” and opacity question “Does Mark know that he found Beth’s treat?”). The DF tasks could also be modified in this way and compared to the AR B and modified KO tasks. The proposed research project seems like a natural progression from tracking the developmental trajectory of children’s understanding of minds in the preschool years through to the early elementary school years. Thus, the unfolding of successful performance on opaque contexts across development remains a question for future research.

Wellman et al.(2001) found that an important conceptual change in children’s understanding of other people’s minds took place between 2;6 and 5 years of age across many cultures. Furthermore, many variations in stimuli and procedures across numerous false belief tasks were irrelevant to successful performance. Although reducing task demands did improve performance, it did not change the developmental trajectory. Even with the effects due to age, task demands, and information processing limitations controlled, the developmental pattern persists. Similar to the false belief tasks then, perhaps minimizing linguistic differences among critical opacity questions across tasks, and creating co-referential terms that strictly fit the definition of opaque contexts will improve performance, but, will it change the developmental trajectory? The answer to this question also awaits future research.

Both categories of tasks (false belief and opacity) presently fall short of measuring a full adult-like understanding of the mind. The creation of tasks that are capable of measuring when children attain a full adult-like understanding of the mind would make an important contribution to the theory of mind literature.

Conclusions

As the title of this thesis suggests, sorting out the relations between false belief and different types of opacity tasks was the focus of the present study. However, the results were mixed. Children in each of the three age groups performed better on the false belief tasks than on the opacity tasks. Although this finding is not new, it provides additional evidence that the former tasks are easier than the latter. False belief tasks were related only to the two opacity tasks using easy opacity questions (ARA and KO). Furthermore, across opacity tasks only the AR A and KO tasks were related. False belief tasks were shown to measure an early developing elementary metarepresentational ability. In contrast, the easy opacity tasks were shown to measure the elementary metarepresentational ability plus the ability to represent co-referential terms as objects of thought separate from the referent. The oldest children in the present study were 6;11 years old and they did not demonstrate mastery of the easy opacity questions. As described above, determining the age at which children master easy and then hard opacity questions are important goals to be explored.

The broad goal of the present study was to further our knowledge about how children come to understand mental events by examining children's performance on tasks that investigate and measure the nature of children's abilities to represent other people's beliefs and knowledge. In the literature, hundreds of studies spanning more than twenty

years have been conducted using the false belief paradigm to measure children's metarepresentational ability. The passing of false belief tasks around the age of four is considered by many to be the hallmark of children's understanding of the mind. Clearly, the few researchers who have recently used opaque contexts to measure children's understanding of the mind found that children's ability to understand mental events develops beyond age four (Apperly & Robinson, 1998; deVilliers & Fitneva, 1996; Kamawar & Olson, 1999; Kamawar, 2000). However, the development of tasks to measure children's understanding of the mind beyond the age of four is in its infancy. The use of opaque contexts is a first step towards measuring later developing understanding of the mind that depends partly on the use of language to denote mental events. However, much work needs to be done before the relations are sorted out among tasks thought to measure how children understand mental events. As described earlier, careful consideration of the critical opacity questions may lead to the development of tasks that become a standard or reliable measure of later developing understanding of the mind just as false belief tasks have become the standard measure of early developing metarepresentational ability. Determining what tasks provide the best measure of opacity remains to be determined in future research.

Creating tasks to effectively measure the later developing understanding of the mind has important implications for children's social cognitive development. Dunn and Cutting (1999) found that among other factors (e.g., emotionality, mother's education) theory of mind skills were related to how well preschool children communicated with their friends. Furthermore, they found that failures in communication with friends were more closely related to theory of mind skills than to vocabulary or emotional

understanding. The link between successful social interactions and children's theory of mind skills implies that understanding the mind is a core social developmental concept. Further support for this notion comes from research on referential opacity with autistic children. Mitchell and Isaacs (1994) note that autistic children have a poor conception of the mind and as such they are poor communicators. They also contend that "when the child enters the realm of unrelated peers ..she must become wise to the cognitions of others if she is to survive in such a context" (p. 450). Hence, understanding the mind may be central to children's success at school and in life. Experiments in the area of opaque contexts can contribute specifically to the advancement of theory-of-mind literature, to the cognitive development literature more generally, and to social skills training programs in schools.

References

- Apperly I. A., & Robinson, E. J. (1998). Children's mental representation of referential relations. *Cognition*, 67, 287-309.
- Apperly I. A., & Robinson, E. J. (2002). Children's handling of referential opacity in relation to their understanding of beliefs as representations. Unpublished manuscript. University of Birmingham.
- Astington, J. W. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.
- Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. *Child Development*, 72 (3), 685-687.
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35 (5), 1311-1320.
- Baddeley, A. D. & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). New York: Cambridge University Press.
- Chandler, J. J., & Sokol, B. W. (1999). Representation once removed: Children's developing conceptions of representational life. In I. E. Sigel (Ed.), *Development of mental representation: Theories and applications* (pp. 201 - 230). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- de Villiers, J. (1994). Questioning minds and answering machines. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the 19th Annual Boston University Conference on Language Development* (pp. 20–36). Somerville, MA: Cascadilla Press.
- de Villiers, J., & Fitneva, S. (1996). *Referential transparency for opaque containers*. Unpublished manuscript, Smith College.
- de Villiers, J., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17, 1037-1060.
- Dunn, J., & Cutting, A. L., (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development*, 8 (2), 201 – 219.
- Engle, R. W., Kane, M. J., & Tuholski, S. W., (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). New York: Cambridge University Press.
- Gathercole, S. E. & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology*, 92 (2), 377-390.
- Hughes, C. (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 19, 233-235.
- Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, 32 (1), 70-78.

- Kamawar, D. (2000). Children's Understanding of the Opaque and Transparent Uses of Language. Unpublished doctoral dissertation. University of Toronto, Canada.
- Kamawar, D., & Homer, B. (2000). Internal and external notions of metarepresentation: A developmental perspective. In J. W. Astington (Ed.), *Minds in the Making*. (pp. 197-211). Blackwell Publishers.
- Kamawar, D., & Olson, D. (1999). Children's representational theory of language: The problem of opaque contexts. *Cognitive Development*, 14, 531-548.
- Mitchell, P., & Isaacs, J. E. (1994). Understanding of verbal representation in children with autism: The case of referential opacity. *British Journal of Developmental Psychology*, 12, 439 – 453.
- Perner, J. (1992). Grasping the concept of representation: Its impact on 4-year-olds' theory of mind and beyond. *Human Development*, 35, 146-155.
- 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, 1451-1472.
- Russell, J., (1987). "Can we say...?" Children's understanding of intentionality. *Cognition*, 25, 289-308.
- Siegler, R. S. (1998). *Children's Thinking*. Upper Saddle River, N. J.: Prentice Hall, Inc.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655-684.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128

Appendix A

Belief Tasks

True Belief Task

Children are presented with a closed bandaids box and asked what is inside:

Look at this... What's in here? [Answer: bandaids]

Let's open it and have a look. What are they? [Correct: bandaids]

Some brief conversation about bandaids... Then... Let's play a trick on X (friend). Let's take the bandaids out and put pennies in here. [Do so and close box]

(When box is closed) What's in the box? [Correct: pennies]

True belief memory question:

What did you think was inside the box BEFORE we opened it? [Correct: bandaids]
This is a screening task and no score was given.

False belief Tasks

1. Change of location task. Cake Story (deVilliers & Fitneva, 1996):

(Picture one) This boy Bobby and his Mom bought a nice cake for after dinner.

(Picture two) But Bobby wanted to go out to play so he put the cake away until after dinner. He put it in **this cupboard** for later. Then he went out to play.

(Picture three) While he was out, the Mom thought that the frosting would melt, so she took the cake out of the cupboard and put it **in the refrigerator**. Then he went out to get some tomatoes for dinner.

Check questions:

Memory: Where did Bobby put the cake?

Reality: Where is it now?

Now Bobby is tired of playing and he's coming home. He remembers where he put the cake.

False belief question: “When he comes in the kitchen, where will Bobby look for the cake?” (false belief question for self – 1 point)

2. Deceptive contents task. Apperly & Robinson (1998): Teddy’s Picture Box Task
Acted out with Stephen the puppet:

Children are shown a box with a picture of a teddy on the outside and asked what they think is inside. After agreeing upon “teddy bear”, the children are allowed to look inside the box and find a horse. After closing the box, children are asked:

Reality question: So, what’s in the box? (Correct = teddy bear = 0 points)

False belief self question: What did you think was in the box *before* we opened it? (Correct = teddy bear = 1 point)

False belief other question: “Stephen hasn’t seen inside this box before. When he first sees it, before he opens it what will he think is inside?” (Correct = teddy bear = 1 point)

3. Deceptive contents task. Kamawar (1999): Crayola Crayon Box Task

Children are presented with a closed Crayola crayon box and asked what is inside:

“Look at this... What’s in here?” [Usual answer: crayons/colours]

“Let’s open it and have a look. What is it?” [Correct answer: stickers]

Some brief conversation about stickers, e.g. “yes! Stickers!
I just put them in this box to keep them safe..... Well... let’s put them back into the box...”

(When box is closed) **Reality question:** “What’s in the box?” (Correct = stickers = 0 points)

***False belief question for self:**

What did you think was inside the box BEFORE we opened it? (Correct = what child said for first question = 1 point)

***False belief question for other:** X (friend’s name) hasn’t seen inside this box.

What will s/he think is inside it before s/he opens it? (Correct = crayons/colours = 1 point).

Appendix B

Referential Opacity Tasks

DeVilliers & Fitneva Tasks Version A:

Trial 1) **Silver Box Story:** I am going to tell you a story about a girl named Amy. As we go along I'm going to ask you some questions about what happens. So, listen carefully, okay?"

(Picture one) This silver box is a birthday present. This birthday present is candy. The Mom put the silver box on the top shelf.

"Did the mom put the silver box on the top shelf?" (Correct = yes)

"Did the mom put the birthday present on the top shelf?" (Correct = yes)

(Picture two) Amy walked into the room one day and saw a silver box on the top shelf. She thought, "I wonder what that silver box is?"

"Does Amy know the silver box is on the top shelf?" (Correct = yes)

"Does Amy know the candy is on the top shelf?" (Correct = no)

She went to her Mom and said, "Mom, why is there a silver box on the shelf!"

(Picture three) Her Mom said, "Do not touch the silver box!"

(Picture four) But Amy took the silver box and put it on the table.

"Did Amy put the silver box on the table?" (Correct = yes)

"Did Amy put the birthday present on the table?" (Correct = yes)

(Picture five) The Mom came in and saw the silver box on the table.

"Does the mom know the silver box is on the table?" (Correct = yes)

"Does the mom know the candy is on the table?" (Correct = yes)

Now the Mom is going to wrap the present and get Amy to help....

Children had to get all the questions right to earn a score of 1 on each of three stories.

Two more similar stories were created in this format (total = 3 points). See below.

Trial 2) **Black Bag Story:** I am going to tell you a story...

(Picture one) This black bag is a special surprise. This special surprise is marbles. The Dad put the black bag in the desk drawer.

“Did the Dad put the black bag in the desk drawer?” (Correct = yes)

“Did the Dad put the special surprise in the desk drawer?” (Correct = yes)

(Picture two) Amy was looking for a pencil one day and saw a black bag in the desk drawer. She thought, "I wonder what that black bag is?"

“Does Amy know the black bag is in the desk drawer?” (Correct = yes)

“Does Amy know the marbles are in the desk drawer?” (Correct = no)

She went to her Dad and said, "Dad, why is there a black bag in the desk drawer!"

(Picture three) Her Dad said, "Do not touch the black bag!"

(Picture four) But Amy took the black bag and put it on the desk.

“Did Amy put the black bag on the desk?” (Correct = yes)

“Did Amy put the special surprise on the desk?” (Correct = yes)

(Picture five) The Dad came in and saw the black bag on the desk.

“Does the Dad know the black bag is on the desk?” (Correct = yes)

“Does the Dad know the marbles are on the desk?” (Correct = yes)

Now the Dad is going to tell Amy about the special surprise and maybe teach her how to play a game of marbles...

Trial 3) **Red Tin Story:** I am going to tell you a story...

(Picture one) This red tin is a Father's day gift. This Father's day gift is cookies. The Mom put the red tin on the counter.

"Did the mom put the red tin on the counter?" (Correct = yes)

"Did the mom put the Father's day gift on the counter?" (Correct = yes)

(Picture two) Amy was colouring one day and saw a red tin on the counter. She thought, "I wonder what that red tin is?"

"Does Amy know the red tin is on the counter?" (Correct = yes)

"Does Amy know the cookies are on the counter?" (Correct = no)

She went to her Mom and said, "Mom, why is there a red tin on the counter!"

(Picture three) Her Mom said, "Do not touch the red tin!"

(Picture four) But Amy took the red tin and put it on the kitchen table.

"Did Amy put the red tin on the kitchen table?" (Correct = yes)

"Did Amy put the Father's day gift on the kitchen table?" (Correct = yes)

(Picture five) The Mom came in and saw the red tin on the kitchen table.

"Does the mom know the red tin is on the kitchen table?" (Correct = yes)

"Does the mom know the cookies are on the kitchen table?" (Correct = yes)

Now the Mom will get Amy to help her wrap the gift...

DeVilliers & Fitneva Tasks Version B

Trial 1) Silver Box Story:

(Picture one) This silver box is a birthday present. This birthday present is candy.
The Mom put the silver box on the top shelf.

Did the mom put the silver box on the top shelf? (Correct = yes)

Did the mom put the candy on the top shelf? (Correct = yes)

(Picture two) Amy walked into the room one day and saw a silver box on the top shelf.
She thought, "I wonder what that silver box is?"

Does Amy know the silver box is on the top shelf? (Correct = yes)

Does Amy know the birthday present is on the top shelf? (Correct = no)

She went to her Mom and said, "Mom, why is there a silver box on the shelf!"

(Picture three) Her Mom said, "Do not touch the silver box!"

(Picture four) But Amy took the silver box and put it on the table.

Did Amy put the silver box on the table? (Correct = yes)

Did Amy put the candy on the table? (Correct = yes)

(Picture five) The Mom came in and saw the silver box on the table.

Does the mom know the silver box is on the table? (Correct = yes)

Does the mom know the birthday present is on the table? (Correct = yes)

Now the Mom is going to wrap the present and get Amy to help....

Trial 2) **Black Bag Story:** *I am going to tell you a story...*

(Picture one) This black bag is a special surprise. This special surprise is marbles. The Dad put the black bag in the desk drawer.

“Did the Dad put the black bag in the desk drawer?” (Correct = yes)

“Did the Dad put the marbles in the desk drawer?” (Correct = yes)

(Picture two) Amy was looking for a pencil one day and saw a black bag in the desk drawer. She thought, "I wonder what that black bag is?"

“Does Amy know the black bag is in the desk drawer?” (Correct = yes)

“Does Amy know the special surprise is in the desk drawer?” (Correct = no)

She went to her Dad and said, "Dad, why is there a black bag in the desk drawer!"

(Picture three) Her Dad said, "Do not touch the black bag!"

(Picture four) But Amy took the black bag and put it on the desk.

“Did Amy put the black bag on the desk?” (Correct = yes)

“Did Amy put the marbles on the desk?” (Correct = yes)

(Picture five) The Dad came in and saw the black bag on the desk.

“Does the Dad know the black bag is on the desk?” (Correct = yes)

“Does the Dad know the special surprise is on the desk?” (Correct = yes)

Now the Dad is going to tell Amy about the special surprise and maybe teach her how to play a game of marbles...

Trial 3) **Red Tin Story:** I am going to tell you a story...

(Picture one) This red tin is a Father's day gift. This Father's day gift is cookies. The Mom put the red tin on the counter.

"Did the mom put the red tin on the counter?" (Correct = yes)

"Did the mom put the cookies on the counter?" (Correct = yes)

(Picture two) Amy was colouring one day and saw a red tin on the counter. She thought, "I wonder what that red tin is?"

"Does Amy know the red tin is on the counter?" (Correct = yes)

"Does Amy know the Father's day gift is on the counter?" (Correct = no)

She went to her Mom and said, "Mom, why is there a red tin on the counter!"

(Picture three) Her Mom said, "Do not touch the red tin!"

(Picture four) But Amy took the red tin and put it on the kitchen table.

"Did Amy put the red tin on the kitchen table?" (Correct = yes)

"Did Amy put the cookies on the kitchen table?" (Correct = yes)

(Picture five) The Mom came in and saw the red tin on the kitchen table.

"Does the mom know the red tin is on the kitchen table?" (Correct = yes)

"Does the mom know the Father's day gift is on the kitchen table?" (Correct = yes).

Now the Mom will get Amy to help her wrap the gift...

Apperly and Robinson Tasks:

Control Task: Duck/Toy Story: acted out with a puppet (Stephen)

Child is asked to look inside the box:

“So, what’s in the box?” Child’s Answer _____.

“Can Stephen see the toy inside the box?” (Correct = yes)

The box is tipped away from the child and the child is asked:

“Can you see the toy with your eyes right now?” (Correct = no)

With the box closed, all children are asked:

“Does Stephen know there’s a toy inside the box” (Correct = yes)

“Do you know there’s a toy inside the box?” (Correct = yes)

No score is given for the control task.

Trial 1). **Dice/Eraser Story:** Child is asked to look inside the box:

“So, what’s in the box?” Child’s Answer _____.

“Go ahead, pick it up and tell me what it feels like.” Child’s Answer _____.

Then: “So it looks like a dice and feels like an eraser.”

“Now Stephen is going to **look** inside the box but he’s not going to **feel**”.

With the box still **open**, the children are asked Q1 (transparent):

“Can Stephen see the (eraser) inside the box?” (Correct = yes)

After the box is **closed** the child is asked Q2 (easy opacity question):

“Does Stephen know the dice is an eraser?” (Correct = no)

* AR version B is identical except the last question is the hard opacity question:

“Does Stephen know there is an eraser in the box?” (Correct = no)

Trial 2). **Ball/Present Story:** Child is asked to look inside the box:

“So, what’s in the box?” Child’s Answer _____.

The experimenter whispers that:

“This is going to be a present for Stephen, except we haven’t told him and we don’t want him to find out right now, so we’ll have to whisper very quietly when he looks...”

Next the children observe as Stephen looks inside the box.

With the box still open, the children are asked Q1 (transparent):

“Can Stephen see the (present) inside the box?” (Correct = yes)

After the box is **closed** the children are asked Q2 (easy opacity question):

“Does Stephen know the ball is a present?” (Correct = no)

* AR version B is identical except the last question is the hard opacity question:

“Does Stephen know there is a present in the box?” (Correct = no)

Trial 3 Car/Surprise Story: Child is asked to look inside the box:

“So, what’s in the box?” Child’s Answer _____.

The experimenter whispers that:

“This is going to be a surprise for Stephen, except we haven’t told him and we don’t want him to find out right now, so we’ll have to whisper very quietly when he looks...”

Next the children observe as Stephen looks inside the box.

With the box still **open**, the children are asked Q1 (transparent):

“Can Stephen see the surprise inside the box?” (Correct = yes)

After the box is **closed** the children are asked Q2 (easy opacity question):

“Does Stephen know the car is a surprise?” (Correct = no)

* AR version B is identical except the last question is the hard opacity question:

“Does Stephen know there is a surprise in the box?” (Correct = no)

Children must answer both questions correctly to earn 1 point for each of three stories.

Kamawar Tasks (not from Kamawar & Olson, 1999, but they demonstrate the same relation to false belief tasks).

Trial1) **Candy/Treat Story**

General introduction to the story: “I am going to tell you a story about a girl named Beth and a boy named Mark. As we go along I’m going to ask you some questions about them, and what they did. So, listen carefully, okay?”

(Picture one) One day, Beth’s mom put some candy in Beth’s lunch bag for her. Beth took her lunch bag to school and put it in her **locker**. There was a hole in her lunch bag, and the candy fell out onto the floor.

(Picture two) Mark walked by and saw the candy on the floor near his cubby. Now, Mark can’t tell that the candy is Beth’s treat. Mark decided that he would like to eat the candy and see what it was like.

(Picture three) Mark walked over to the candy, and picked it up. He noticed that the wrapper was really AR B to get off. He pulled at it and it wouldn’t come off. He said, “I’ll use scissors to cut the wrapper off the candy.”

(Picture four) Mark got some scissors and cut the wrapper off the candy.

(Picture five) Mark threw the wrapper away and put the candy in his mouth. He smiled a big smile and said, “This is really good”.

Questions:

- 1) Did Mark find an apple? (Correct answer = no)
- 2) Did Mark find a candy? (Correct answer = yes)
- 3) Did Mark find Beth’s treat? (Correct answer = yes)
- 4) Does Mark **know** that he found a candy? (Correct answer = yes)

5) Does Mark **know** that he found Beth's treat? (Correct answer = no)

Really, if you find a candy on the floor you should tell the teacher and not just go ahead and eat it...

Children had to answer all questions correctly to earn 1 point for each of three stories.

Trial 2). **Truck/Present Story**

(Picture one) One day, Mark was playing in his parents' room and he opened up their closet door. Up high, on the top shelf was a new, shiny, red truck. The truck is going to be a birthday present for Mark's cousin- Mark's parents were keeping it in their closet until the cousin's birthday. Now, Mark can't tell that the present is for his cousin. Mark decided that he wanted to play with the truck. He reaches up high to get the truck, but he can't reach it because it is way too high. He says out loud, "I am going to knock down the truck."

(Picture two) He thinks, and then he says, "I know, I'll use my baseball to knock down the truck." So, Mark threw his baseball up and he missed the truck."

He tried again, and this time it worked. The baseball hit the truck and knocked it down.

Mark said, "I knocked down the truck!"

(Picture four) Then, Mark played with the truck in his parent's room.

Questions:

1) Did Mark throw a ball at a doll? (Correct answer = no)

2) Did Mark throw a ball at a birthday present? (Correct answer = yes)

3) Did Mark throw a ball at a truck? (Correct answer = yes)

4) Does Mark **know** that he threw a ball at a truck? (Correct answer = yes)

5) Does Mark **know** that he threw a ball at a birthday present? (Correct answer = no)

Marks parents will have to tell him who the truck is for...

Trial 3. **Money/Mark's Allowance Story**

(Picture one) After school one day, on his way to the store, Mark stopped because he saw that his shoelaces were undone. He was holding some money in his hand, and had to put it down because he needed both hands to tie his shoe laces.

(Picture two) After he was done, he forgot to pick up the money and went on his way.

(Picture three) A little later, Beth walked by. She saw some money on the ground. Now, Beth couldn't tell that the money was Mark's allowance. Beth decided she would take the money and spend it.

(Picture four) Beth walked over to the chip machine. She decided to get some potato chips. She reached up to put the money in the machine and missed. She said, "I'll get a crate so I can reach".

(Picture five) and she went and got a crate that was nearby. When Beth stood on the crate, she could reach and put the money in the machine and got her chips. Then she said, "I'm glad I found that money."

Questions:

- 1). Did Beth pick up her Mom's money? (Correct answer = no)
- 2). Did Beth pick up Mark's allowance? (Correct answer = yes)
- 3). Did Beth pick up some money? (Correct answer = yes)
- 4) Does Beth **know** that she picked up some money? (Correct answer = yes)
- 5). Does Beth **know** that she picked up Mark's allowance? (Correct answer = no)

Really, if you find money on the ground you should tell a grown up like your parents or your teacher and not just take it and spend it...

* For trials 2 and 3 the general story introduction begins with “Remember, I am going to tell you a story about two friends, Mark and Beth...”

Appendix C

Complement Comprehension

Materials: Binders with pictures

Protocol: Read the story pointing to the important objects in each pictures. Point again to the first picture when asking the question.

What-think (other)

She thought the girl was reading a book, but she was really playing cards.

13.) *What did she think?* _____

What-tell (other)

She told the girl there was a bug in her hair, but it was only a leaf.

14.) *What did she tell the girl?* _____

What-tell (self)

She told her husband she saw a ghost, but it was really a blanket.

15.) *What did she tell her husband?* _____

What-think (self)

She thought she had a hole in her pants, but it was really a piece of paper.

16.) *What did she think?* _____

What-tell (other)

She told her dad he had a cut, but it was really ketchup.

17.) *What did she tell her dad?* _____

What-tell (self)

She told the teacher she drew a face, but it was really a scribble.

18.) *What did she tell the teacher?* _____

What-think (other)

She thought her friend was eating an egg, but it was really a ball.

19.) *What did she think?* _____

What-think (self)

She thought there was a spider in her cereal, but it was really a berry.

19.) *What did she think?* _____

* The complex syntax (i.e., complementation comprehension) test is not a standardized test. It has previously been used by de Villiers and Pyers (2002) in children aged 3-5 years as one test among a battery of tests to assess false belief understanding and language mastery. They found that mastery of tensed complements is a precursor and possibly a prerequisite of successful false belief performance.

Appendix D

Digit Span Tasks

Digit Span: Backward

“Now I am going to say some more numbers. This time I want you to say them backwards. For example, if I say 5 – 8, you would say 8 – 5. Do you understand? Let’s try one. Ready? Listen carefully. Remember to say the numbers backwards, 2 – 4”.

If correct, proceed. Else, say “You should say 4 - 2. Remember, I said 2 –4. To say 2 – 4 backwards, you would say 4 – 2”

Proceed to sample item 2.

“Let’s try another one. Ready? Listen carefully. Remember to say the numbers backwards, 7 - 1”

If correct, proceed. Else, say “You should say 1 – 7. Remember, I said 7 – 1. To say 7 – 1 backwards, you would say 1 - 7”.

For each Digits Backwards item, say the digits at the rate of one per second.

Item:	Digits Forward:	Child’s Response:	P/F:
S1.	2 – 4	4□ – 2□	P / F
S2.	7 – 1	1□ – 7□	P / F
1.	6 – 3	3□ – 6□	P / F
2.	4 – 9	9□ – 4□	P / F
3.	2 – 9 – 5	5□ – 9□ – 2□	P / F
4.	8 – 1 – 6	6□ – 1□ – 8□	P / F
5.	8 – 5 – 2 – 6	6□ – 2□ – 5□ – 8□	P / F
6.	4 – 9 – 3 – 7	7□ – 3□ – 9□ – 4□	P / F
7.	8 – 1 – 3 – 7 – 9	9□ – 7□ – 3□ – 1□ – 8□	P / F
8.	4 – 2 – 5 – 8 – 1	1□ – 8□ – 5□ – 2□ – 4□	P / F
9.	4 – 8 – 7 – 2 – 1 – 5	5□ – 1□ – 2□ – 7□ – 8□ – 4□	P / F
10.	6 – 2 – 5 – 9 – 3 – 8	8□ – 3□ – 9□ – 5□ – 2□ – 6□	P / F
11.	1 – 8 – 4 – 2 – 5 – 9 – 3	3□ – 9□ – 5□ – 2□ – 4□ – 8□ – 1□	P / F
12.	4 – 7 – 3 – 9 – 6 – 1 – 2	2□ – 1□ – 6□ – 9□ – 3□ – 7□ – 4□	P / F

Highest item number administered =

Total number of failed items = _____

Raw Score =

Digit Span: Forward

"I am going to say some numbers. Listen carefully and try to say them just as I do. Ready? Listen carefully, 5 – 8".

If correct, proceed. Else, say "You should say 5 – 8. That's what I said"

Proceed to sample item 2.

"Let's try another one. Ready? Listen carefully. Say the numbers just as I do, 3 – 9"

If correct, proceed. Else, say "You should say 3 – 9. That's what I said"

For each Digits Forward item, say the digits at the rate of one per second.

Item:	Digits Forward:	Child's Response:	P/F:
S1.	5 – 8	5□ – 8□	P / F
S2.	3 – 9	3□ – 9□	P / F
1.	5 – 7 – 8	5□ – 7□ – 8□	P / F
2.	4 – 9 – 2	4□ – 9□ – 2□	P / F
3.	2 – 7 – 6 – 9	2□ – 7□ – 6□ – 9□	P / F
4.	5 – 1 – 8 – 4	5□ – 1□ – 8□ – 4□	P / F
5.	3 – 1 – 8 – 5 – 9	3□ – 1□ – 8□ – 5□ – 9□	P / F
6.	4 – 8 – 3 – 7 – 2	4□ – 8□ – 3□ – 7□ – 2□	P / F
7.	2 – 8 – 3 – 5 – 9 – 4	2□ – 8□ – 3□ – 5□ – 9□ – 4□	P / F
8.	7 – 1 – 9 – 5 – 4 – 3	7□ – 1□ – 9□ – 5□ – 4□ – 3□	P / F
9.	3 – 5 – 9 – 6 – 8 – 4 – 7	3□ – 5□ – 9□ – 6□ – 8□ – 4□ – 7□	P / F
10.	2 – 8 – 5 – 1 – 4 – 6 – 9	2□ – 8□ – 5□ – 1□ – 4□ – 6□ – 9□	P / F
11.	7 – 3 – 9 – 6 – 8 – 1 – 4 – 2	7□ – 3□ – 9□ – 6□ – 8□ – 1□ – 4□ – 2□	P / F
12.	2 – 8 – 1 – 7 – 3 – 5 – 9 – 6	2□ – 8□ – 1□ – 7□ – 3□ – 5□ – 9□ – 6□	P / F
13.	8 – 1 – 7 – 4 – 9 – 3 – 2 – 6 – 5	8□ – 1□ – 7□ – 4□ – 9□ – 3□ – 2□ – 6□ – 5□	P / F
14.	5 – 7 – 4 – 2 – 6 – 1 – 3 – 9 – 8	5□ – 7□ – 4□ – 2□ – 6□ – 1□ – 3□ – 9□ – 8□	P / F

Highest item number administered =

Total number of failed items =

Raw Score =

Appendix E

Information Letter and Consent Form

Dear parent(s) or guardian(s),

We are conducting a study of cognitive development in children aged 3 yrs. 9 mos. to 6 yrs. 11 mos. and would like to include your child in our study. The study has been approved by the Carleton University Ethics Committee for Psychological Research and will be conducted in accordance with their ethical standards; it involves no physical or psychological risks. The Ottawa-Carleton Research Advisory Committee has approved the study for implementation in the School Board. The purpose of this informed consent is to ensure that you understand the goals of the study and the nature of your child's involvement.

The study will examine the development of children's ability to represent what other people believe and know. For example, in one type of story, a boy puts a cake in a cupboard. While the boy is out playing, the Mom moves the cake to the refrigerator so the frosting won't melt. The children's task is to answer the question "Where will the boy first look for the chocolate when he comes back into the kitchen?" In another type of story, children must keep track of what the story character knows about the specific details of an object. For example, a little boy finds a toy truck, but doesn't know that the truck is going to be a birthday gift for his cousin. Children are asked questions about what the boy is thinking about in relation to the toy (e.g., "Does he know that he's playing with a truck?" or "Does he know that he's playing with a birthday present?"). The children will also complete two short tasks; one that looks at general memory ability and one that evaluates understanding of a particular aspect of grammar. Please note that we are not assessing individual children but examining performance on different tasks across different ages in order to increase our knowledge of children's development in general.

The research will be conducted at your child's school. Your child's participation will last approximately 10-15 minutes for each of three sessions. To ensure the experimenter's accuracy in recording the children's responses, sessions will be audiotaped (voice only). All data collected in this study are confidential and are coded such that the child's name is not associated with the data. The coded data are made available only to the researchers associated with this project.

In the past, children have generally enjoyed taking part in these kinds of tasks. Children will be given stickers to thank them for their participation. The researchers will be sensitive to the children and *all* answers will be met with a positive response. Participating children may withdraw from the study at any time (and still receive their stickers).

If you have any questions about the research, please contact either Dawn Mullins, who may be reached at 520-2600, ext.2885, or the faculty sponsor, Dr. Deepthi Kamawar (520-2600, ext 7021). If you have any concerns about how this study was conducted, please contact Dr. Mary Gick (Chair, Carleton University Ethics Committee for Psychological Research, 520-2600 ext. 2664) or Dr. John Logan (Chair, Dept. of Psychology, 520-2600 ext. 2648).

Kindly sign the attached consent form indicating whether your child may participate in this research. If you would like a summary of the research results once the study is completed, please include your mailing address on the attached form. Thank you for your attention.

Sincerely,
Dawn B. Mullins

Parent's Signature

I have read the above description of the study and I understand the conditions of my child's participation. My signature indicates that I agree to let my child participate in the study.

Child's Name: _____ Child's Date of Birth: _____

Parent's Signature: _____ Date: _____

Child's Teacher: _____

Include your mailing address here if you wish to receive a copy of the study later.

Mailing address:

Appendix F

Experiment Design

	FB Level 1	Opacity Level 1	Opacity Level 2	Opacity Level 3	Opacity Level 3
Age Group	FB	DF (Transparent Q and Opaque Q)	KO (Transparent Q and Opaque Q)	AR	AR
Level 1 Young (3;9 – 4;9)	S1 S2 S.. S20	S1 S2 S.. S20	S1 S2 S.. S20	AR A (Transparent Q and Opaque Q-A) S1 S2 S.. S10	AR B (Transparent Q and Opaque Q-B) S31 S32 S.. S40
Level 2 Middle (4;10 – 5;10)	S21 S22 S.. S40	S21 S22 S.. S40	S21 S22 S.. S40	AR A (Transparent Q and Opaque Q-A) S11 S12 S.. S20	AR B (Transparent Q and Opaque Q-B) S41 S42 S.. S50
Level 3 Old 5;11 -6 11)	S41 S42 S.. S60	S41 S42 S.. S60	S41 S42 S.. S60	AR A (Transparent Q and Opaque Q-A) S21 S22 S.. S30	AR B (Transparent Q and Opaque Q-B) S51 S52 S.. S60

All 60 children received FB, DF, and KO

FB (False Belief) has 1 level and is a different variable from opacity

Opacity has 3 levels (DF, KO, and AR)

All 60 children received two important types of questions (transparent and opaque) at the DF level and KO level of opacity (20 children in each of the 3 age groups = 60)

At the AR levels of opacity children received two important types of questions (transparent and opaque but there were two types of opaque questions):

ARA --30 children got a transparent question and an easy opaque question (10 children in each of the 3 age groups = 30) in AR "A"

ARB--30 children got a transparent question and a hard opaque question (10 children in each of the 3 age groups = 30) in AR "B"

Appendix G

Comparisons of Critical Questions across Opacity Tasks

	Transparent	Opacity	Opacity
DF	Did the mom put the silver box on the top shelf? ($\checkmark$ = yes) Did the mom put the birthday present/candy on the top shelf? ($\checkmark$ = yes) Did Amy put the silver box on the table? ($\checkmark$ = yes) Did Amy put the birthday present on the table? ($\checkmark$ = yes)	Does Amy know the silver box is on the top shelf? ($\checkmark$ = yes) Does Amy know the candy/birthday present is on the top shelf? ($\checkmark$ = no) Does the mom know the silver box is on the table? ($\checkmark$ = yes) Does the mom know the candy/birthday present is on the table? ($\checkmark$ = yes)	
A R	Can Stephen see the present inside the box? ($\checkmark$ = yes).	AR A (Easy) Does Stephen know the ball is a present? ($\checkmark$ = no).	AR B (Hard) Does Stephen know there's a (present) in the box? ($\checkmark$ = no).
K O	Did Mark find an apple? ($\checkmark$ = no) Did Mark find a candy? ($\checkmark$ = yes) Did Mark find Beth's treat? ($\checkmark$ = yes)	Does Mark know that he found a candy? ($\checkmark$ = yes) Does Mark know that he found Beth's treat? ($\checkmark$ = no)	