In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Psychology**

The University of British Columbia
Vancouver, Canada

Date **5 Nov 96**
ABSTRACT

Six studies were conducted to investigate young children's earliest insights into the interpretive nature of knowing, or the formation of what has been called an “Interpretive Theory of Mind.” This insight was operationalized as the ability to recognize that two persons exposed to precisely the same information can, nonetheless, end up holding sharply different opinions about what is the self-same reality. All of the studies made use of a set of pictures fitted with covers such that most of the underlying picture was hidden, leaving only an ambiguous set of lines visible through a small viewing window. The key question asked concerned subjects' understanding that other persons who have not seen the full picture but only the restricted view, and who know nothing about the full contents of the picture, are all nevertheless free and able to hold different beliefs about what is depicted in this restricted view. An important feature of this procedure is that it can assess both subjects' understanding of simple false belief as well as their understanding of the interpretive possibilities that such stimuli afford. This feature was exploited to demonstrate that young persons who appreciate that beliefs can be false—an achievement that is commonly taken to mark the point of entry into a theory-like understanding of mental life—cannot always be counted on to also appreciate that different interpretations of the same stimulus are possible. By exploring children's reactions to inherently ambiguous stimuli that, by design, easily lend themselves to misinterpretation, it is possible to distinguish between a theory of mind that rests entirely on notions of false belief (i.e., one that views the mind as a recording device capable only of mistakes of ignorance), and a more complex appreciation of the mind's more active capacity for constructively interpreting—and so misinterpreting—reality. What these studies reveal is that an interpretive theory of mind is different from, and later arriving than, an appreciation of the possibility of false belief, and, contrary to competing claims, this interpretive theory actually makes its first appearance during, but not before, the early school years.
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii
Table of Contents ................................................................................................................... iii
List of Tables .......................................................................................................................... v
List of Figures ......................................................................................................................... vii
Acknowledgments ................................................................................................................ .. viii
Introduction .............................................................................................................................. 1
   Statement of the Problem ..................................................................................................... 1
Review of the Relevant Literature ......................................................................................... 4
   False Belief and the Mind as a Representational Medium ................................................ 5
   Evidence of Children’s Understanding of False Belief ....................................................... 7
   Equating Interpretation with False Belief .......................................................................... 10
Narrowing Upon an Appropriate Set of Assessment Tools ................................................ 15
   Promising Leads ................................................................................................................. 16
      Perspective Taking Skills ............................................................................................... 16
      Role-Taking Skills ......................................................................................................... 17
      Conceptual Perspective Taking Skills ........................................................................... 21
      The Mind as an Active Information Processor ............................................................. 29
      Biased Beliefs .................................................................................................................. 31
      Matters of Taste, Ambiguity, and Interpretation ............................................................ 33
A Proposed Methodology ...................................................................................................... 38
The Experiments .................................................................................................................... 44
Study 1 .................................................................................................................................... 46
   Subjects ............................................................................................................................... 48
   Materials ............................................................................................................................. 48
   Design & Procedure .......................................................................................................... 48
   Results ................................................................................................................................. 49
      Classification of Responses ............................................................................................ 49
      False Belief Attributions ............................................................................................... 50
      Interpretive Attributions ............................................................................................... 51
      Stimulus Suitability ....................................................................................................... 51
   Discussion ........................................................................................................................... 52
Study 2 .................................................................................................................................... 54
   Subjects ............................................................................................................................... 55
   Materials ............................................................................................................................. 55
   Design & Procedure .......................................................................................................... 55
   Results ................................................................................................................................. 57
      Scoring of Individual Responses .................................................................................... 57
      Scoring of Response Pairs ............................................................................................. 59
      False Belief Understanding ............................................................................................ 59
      Interpretive Understanding ............................................................................................ 63
      Comparing False Belief and Interpretive Understanding .............................................. 66
   Discussion ........................................................................................................................... 68
<table>
<thead>
<tr>
<th>Study</th>
<th>Pages</th>
<th>Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 3</td>
<td>70</td>
<td>Subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Materials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design &amp; Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Belief Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpretive Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparing False Belief and Interpretive Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>Study 4</td>
<td>85</td>
<td>Subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Materials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design &amp; Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Belief Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpretive Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control Trial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparing False Belief and Interpretive Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>Study 5</td>
<td>96</td>
<td>Subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Materials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design &amp; Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>Study 6</td>
<td>99</td>
<td>Subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Materials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design &amp; Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Belief Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpretive Understanding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use &amp; Recall of Guesses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Justification of Beliefs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>General Discussion</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Bibliography</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>Appendices</td>
<td>121</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Percentage of subjects giving one or fewer Level 1 responses (adapted from Taylor, 1988) ..........................................................24
Table 2: Mean number of correct responses by Age Group and Condition (adapted from Taylor, 1988) ..........................................................27
Table 3: Classification of Responses for False Belief Understanding (Restricted View Trials by Age Group, Study 2) ........................................60
Table 4: False Belief Understanding Scores (Restricted View Trials by Age Group, Study 2) ..........................................................60
Table 5: Classification of Responses for False Belief Understanding (Unrestricted View Trials by Age Group, Study 2) ......................................61
Table 6: False Belief Understanding Scores (Unrestricted View Trials by Age Group, Study 2) ..........................................................62
Table 7: Classification of Responses for Interpretive Understanding (Restricted View Trials by Age Group, Study 2) ........................................63
Table 8: Classification of Individual Subjects' Understanding of Interpretation (Restricted View Trials by Age Group, Study 2) .........................64
Table 9: Classification of Responses for Interpretive Understanding (Unrestricted View Trials by Age Group, Study 2) ........................................64
Table 10: Classification of Individual Subjects' Understanding of Interpretation (Unrestricted View Trials by Age Group, Study 2) ....................65
Table 11: Contingency Table of Subjects' False Belief and Interpretive Classifications (Study 2) ..........................................................67
Table 12: Distribution of subjects across Conditions by Age and Sex (Study 3) ..........................................................72
Table 13: Classification of Responses for False Belief Understanding by Age Group (Study 3) ..........................................................77
Table 14: False Belief Understanding Scores by Age Group (Study 3) ..........................................................77
Table 15: Classification of Individual Subjects' False Belief Understanding by Age Group (Study 3) ..........................................................78
Table 16: Contingency Table of False Belief Classifications (Study 2 and Study 3) ..........................................................79
Table 17: Classification of Responses for Interpretive Understanding by Age Group (Study 3) ..........................................................79
Table 18: Interpretive Understanding Scores by Age Group (Study 3) ..........................................................80
Table 19: Interpretive Understanding by Age Group (Study 3) ..........................................................81
Table 20: Interpretive Understanding by Condition (Study 3) ..........................................................81
Table 21: Changes in Subjects' Interpretive Scores from Study 2 to Study 3 ..........................................................82
Table 22: Changes in Subjects' Interpretive Classifications from Study 2 to Study 3 ..........................................................82
Table 23: Contingency Table of Subjects' False Belief and Interpretive Classifications (Study 3) ..........................................................82
Table 24: False Belief Understanding by Age Group (Study 4) ..........................................................90
LIST OF FIGURES

Figure 1: Droodle labeled “A ship arriving too late to save a drowning witch”.......................18
Figure 2: The “Ship-Witch” droodle.................................................................18
Figure 3: Stimuli used by Taylor (1988, Study 1). ..................................................24
Figure 4: Restricted views used by Taylor (1988, Study 2). ........................................26
Figure 5: Ambiguous figures used by Carpendale (1995).............................................35
Figure 6: Experimental set up for Study 2.............................................................56
Figure 7: Mean False Belief scores by Trial Type and Age Group (Study 2)......................62
Figure 8: Percentage of Interpretive Response Types (Restricted View Trials by Age Group, Study 2). .................................................................63
Figure 9: Percentage of Interpretive Response Types (Unrestricted View Trials by Age Group, Study 2). .................................................................64
Figure 10: Mean Interpretive Scores by Trial Type and Age Group (Study 2).....................65
Figure 11: Mean False Belief and Interpretive Scores by Trial Type (Study 2).....................66
Figure 12: Percentage of Interpretive Response Types by Age Group (Study 3)...................80
Figure 13: Two window version of the ship-witch droodle...........................................87
Figure 14: Interpretive Response Types by Age Group (Study 4) .....................................92
ACKNOWLEDGMENTS

Financial support for this dissertation was provided by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada and in part by a research grant held by Michael Chandler from the Natural Sciences and Engineering Research Council of Canada. It was, I think, money well spent.

A warm thank-you is extended to the children of Spare Time Fun Centre who made up the bulk of the subjects in these studies and to the staff who were enthusiastic proponents of the research. Special thanks also go to the parents and players from the Marpole Little League and Marpole Soccer Club and to my friends and neighbours (and their children) at Vera Housing Cooperative who helped round out the subject pool. They are friends indeed.

Thanks are also due the following supporters:

To Lou Moses who helped frame parts of studies 4 and 6. To Janet Werker for her invaluable service as a member of my dissertation committee and for nearly a dozen years of academic guidance and supportive friendship. To Charlotte Johnston (who cheerfully served on every committee I was obliged to form as a student) for her unflagging ability to bring weighty epistemological issues gently back to earth. To Peter Graf and Geoff Hall for their insightful comments on an earlier draft of this document. To Jeremy Carpendale for his efforts as a co-conspirator in the fight to unseat the “one-miracle” view. To Judith Pegg for her long-standing, gentle yet deadly serious insistence that I “finish it!” To my dear friend Stuart McKenzie (see Appendix IV) who warmly applauded every small triumph and winced at every setback that I experienced as a graduate student.

To my supervisor, Michael Chandler, who, as a result of my hapless enrollment in his section of Psychology 301, is to be held personally responsible for my decision to enter graduate school (and without whose wisdom and guidance this work would not exist), there is simply nothing that I could write that would properly convey your value to me as a mentor and a friend. You raised the sail, I just trimmed it.

To my parents and siblings and in-laws and siblings in-law, thanks for your love and support. To my loving and long-suffering wife Laurie who can uniquely appreciate the significance of the following reference: My mother is right—you are going straight to heaven. And finally, to my children Peter, Riley, and Lise, thank you for your patience in serving as dissertation guinea pigs in a seemingly endless series of Dad’s “games.” Patience in a child is a rare and wondrous thing. But then, you are rare and wondrous children. Daddy’s done now—can we play one of your games?

This work is dedicated with love to Mary (“Granny”) Kehoe and Edith (“G.G”) Henderson.
INTRODUCTION

STATEMENT OF THE PROBLEM

This study sequence was designed to explore the developmental course by means of which young children come to the common adult view that the process of knowledge acquisition is ineluctably interpretive—that knowledge is shaped not only by an external reality but also by the mind that seeks to understand it. A central feature of this interpretive understanding is that multiple interpretations of the same object or event are almost always possible, are usually permissible, and (on some occasions) are all but inevitable. Appreciating the possibility of “interpretive diversity” (Taylor, Cartwright & Bowden, 1991) and working to explain or resolve such differences forms a part of everything from psychotherapy to international relations. To be properly credited with a working grasp of the interpretive character of knowing itself, however, one would, at a minimum, need to appreciate that it is possible for two persons exposed to precisely the same information to nevertheless come to different conclusions, beliefs, or interpretations as to the meaning of that information. While much more will need to be said about the concept of interpretation and about just what it might mean to have different interpretations of one and the same thing, the thesis advanced here is that the developmental debut of this interpretive understanding does not take place until the early school years. The six studies to be reported below were designed to establish this alleged fact, and to help distinguish the achievement of an understanding of interpretation from what is shown to be an altogether different ability that marks the preschooler’s growing appreciation of the role that simple false belief plays in mental life.

At least three testable intuitions underpin the framing of this research programme. The first is that, within our own culture at least, ordinary adults are best understood as more or less explicitly subscribing to an understanding of the knowing process according to which the act of becoming knowledgeable about any new thing is always taken to be a joint product of contributions made by the stimulus environment on the one hand, and by whatever is true about the knowing person on the other. The common sense epistemology or “theory of mind” subscribed to by most adults is of
just this double-barreled variety and proceeds on an understanding of learning that involves what John Searle (1983) calls both a “world to mind” and a “mind to world” direction of fit. By this account, minds need to be seen, not only as fitting themselves to an independent reality, but also as being capable of altering or deforming that reality in such a way as to make it better conform to the particular sorts of intelligence seeking to understand it. Anything that might begin to qualify as a fully fledged “interpretive” or “constructive” theory of mind, would somehow need to make room for the fact that mental contents are never directly copied from the environment, but are always constructed or shaped to a greater or lesser extent by a mind that is irrepressibly active.

The second broad intuition underpinning this research is that young persons are not born into the world already subscribing to anything like an “interpretive” theory of the mind, but instead only gradually come to such an interactive view during the course of their development. More specifically, the claim tested here is that young preschool children ordinarily start out their careers as apprentice theorists of the mind by initially subscribing to only the first half of the whole double-edged truth of this matter. That is, they begin by operating as though minds are only obliged to “fit” the world, that only exogenous factors shape mental life, that people can only passively accommodate to the pressures of outside experiences, and that they always passively copy rather than actively construct a reality that they assume to be given, apart from any creative efforts on their part to come to know it. In short, while a mature view of mental life necessarily presupposes that minds are both “assimilative” and “accommodative,” young preschool children at first fail to appreciate this fact. Instead, they begin their early attempts at psychological “theory” construction by first employing a view of knowledge and belief that turns upon the more limited notion that minds necessarily fit themselves to the world, and persist in holding to this one-sided view for some time before going on to make room for the equally critical counterpart idea of a world-to-mind direction of fit. It is this second achievement, I will maintain, that serves to define those novel accomplishments that prove to set 5- to 7-year-olds apart from their still younger preschool counterparts, and that marks them as having come to a more fully fledged understanding of mental life. This expectation carries with it the obligation, actively taken up in this research, to
come to some reasonably precise understanding of when, and perhaps how, children come to this supposedly late-arriving realization.

Third, and finally, the work carried out for this dissertation holds to empirical test the notion that many of those currently contributing to the study of children’s developing theories of mind are actually mistaken in their shared belief that the emerging capacity of young 3- to 5-year-old subjects to entertain the possibility of false belief amounts to the same thing as (or can be counted as evidence for) what is hypothesized here to be a different and later arriving “interpretive” theory of mind. As a way of pre-empting any ‘straw man’ notions that might be associated with this claim, it should be noted at the outset that the set of researchers who make such claims includes such marquee players as Henry Wellman, Josef Perner, and John Flavell. Wellman, for one, sees the preschool period as a time when children develop an “interpretational or constructive understanding of representations” (Wellman, 1990, p.244). Similarly, Flavell contends that an understanding of false belief is a special case of understanding the “appearance-reality” distinction and that children come to grasp the “one-many” relation that obtains between things in the world and their possible representations during this same period, giving them the ability to see that “the same object can be represented in different, seemingly contradictory ways” (Flavell, 1988, p.247). Josef Perner endorses this view more directly: “around 4 years children begin to understand knowledge as representation, with all its essential characteristics. One such characteristic is interpretation” (Perner, 1991, p. 275, italics original). In contrast to these views, the research reported in this dissertation demonstrates that false-belief understanding is a necessary but not sufficient condition for what appears to be a later arriving ability to appreciate that two or more persons can, and often do, interpret the self-same information differently. In the literature review that follows, these contrasting ‘interpretations’ of the available evidence are examined.
REVIEW OF THE RELEVANT LITERATURE

The main aim of this research is to establish that there exists both a theoretical and a developmental difference between two seemingly similar ways in which adults commonly characterize the mental lives of other persons. The first of these is an everyday understanding that people can hold to, and act upon, beliefs that are objectively false. The second is that people can legitimately hold any of a range of different beliefs about precisely the same particular state of the world. Our naive or “folk psychology” (D'Andrade, 1987) follows from both of these routine, adult, insights into other minds. The first is a representational view that depends on a distinction between the world and mental representations of the world that allows us to infer unobservable mental states in others, and permits us to use those inferences to predict and explain their behaviour (Wellman, 1988). The second is more constructivistic or dialogic and makes room for a wider array of possible ways in which reality might be construed and extends to circumstances in which competing beliefs about the world are not so easily sorted into ‘true’ as opposed to ‘false’ representations.

Most of the research carried out to date in the area of children’s developing theories of mind has centered on the first of these two insights, that is, on children’s understanding of the concept of belief, and more specifically on their appreciation of the notion that beliefs can be false. What this work has succeeded in showing is that at some point between the age of 2 and 4 years, children come to appreciate the mentalistic notion that people typically behave with respect to their beliefs about reality rather than with respect to “reality” itself, and, therefore, to understand that people sometimes act on the basis of beliefs that are objectively false (Chandler, Fritz & Hala, 1989; Freeman, Lewis & Doherty, 1991; Fritz, 1992; Hala, Chandler & Fritz, 1991; Lewis & Osborne, 1990; Mitchell & Lacohee, 1991; Wellman, 1990; Wimmer & Perner, 1983). This understanding is widely held out as signaling the arrival of a theory-like appreciation of the mind as a representational device capable of holding both factual and counterfactual representations of reality. Further, this minimal understanding of false belief is said to be prerequisite to our commonsense view of the social world according to which the actions of others can be predicted.
and explained by making presumptions about their internal mental states—their beliefs, desires, and intentions (Astington, Harris & Olson, 1988; Frye & Moore, 1991; Lewis & Mitchell, 1994; Perner, 1991; Wellman, 1990). The importance of developing such a mentalistic view of others is underscored by recent research showing that many of the central features of autism may be explained as an inability to acquire a theory of mind (Baron-Cohen, Tager-Flusberg, Cohen & Volkmar, 1993).

The purpose of this literature review is: a) to examine what is known about children’s understanding of the mind as a “representational medium” (Perner, 1988a; Perner, 1991); b) to explore why it is that an understanding of false belief in particular has become the litmus test for the acquisition of a theory of mind, and why many in the field seem to believe that the first theory children acquire is, for all intents and purposes, also their last; and finally, c) to go on to describe currently available research that, like the studies in this dissertation, touch on children’s appreciation of the interpretive or constructive capacities of the mind that undergird a broader conception of “interpretive diversity” (Carpendale, 1995; Carpendale & Chandler, in press; Taylor et al., 1991).

**False Belief and the Mind as a Representational Medium**

The common thread that runs through essentially all of the research into children’s developing theories of mind is a shared concern with the developmental course by means of which young children come to an understanding of mind as a representational medium or device. In describing the capacity of the mind to represent states of the world, Josef Perner (1991) has argued that any representation has both a ‘referent’ and a ‘sense.’ According to Perner, the referent is simply that thing in the world that the representation is about—the child that is depicted in the painting of that child. The sense of a representation is its propositional content—the object or event is depicted in some way or some state, for example the child is depicted as ‘innocent.’ As a representational device or medium then, the mind is seen to represent things as being a particular way. As a mental representation, any particular belief, such as my belief that the glass is half empty, has both a
referent (a thing that it can be said to be 'about') as well as a sense (the thing as being a certain way).

Two important points need to be brought out about beliefs as mental representations. The first is that beliefs can misrepresent the thing (or the state of the thing) that they are meant to be about. To claim that a young child (or anyone else for that matter) understands that the mind is capable of forming representations of the world, one would need to be convinced that the child understands the possibility of misrepresentation, since the whole point about beliefs (as Perner and others remind us) is that they can be false. For this reason, research into children’s understanding of mental states has largely concentrated on their appreciation of the possibility of false belief—that you or I might take as true what is objectively a false representation of some state of the world.

There is, in addition, a practical reason for concentrating on false belief that has to do with the fact that responses to questions about the true beliefs of another person cannot be distinguished from simple reports of one’s own belief about the same matter, since these amount to exactly the same thing.

The second important implication that follows from a conceptualization of the mind as a representational medium is that beliefs and other mental states are fundamentally private and unobservable. That is, if left undeclared, the presence and content of such states can only be inferred. When we attempt to predict or explain the behaviour of others, we make inferences about their beliefs, desires and intentions: “he is going to the kitchen because he wants some juice, he will look in the refrigerator because he thinks there is juice there, but he doesn’t know that I drank it all, so he will be disappointed when he sees the empty jug.” Wellman (1988) and others have argued—convincingly I think—that the foundation of this “belief-desire psychology” is a theory-like understanding of the mind. This “theory-theory” (Morton, 1980), as it has come to be called, holds that children’s early appreciation of mental states and the relations between such states and behaviour can be seen to form a proper—if fledgling—theory. As Wellman states, it is “a naive theory, not a developed or disciplined scientific theory, but a theory nonetheless” (1988, p.67). To back such strong claims, Wellman points to children’s appropriate use of mental state terms and to
their ability to accurately predict and explain the actions of others. This evinces an understanding of mental life in which “notions such as thoughts, dreams, beliefs, desires, and intents—are an interconnected coherent body of concepts including theoretical terms; they rest on and define basic ontological conceptions; and the theory provides a causal-explanatory account of a relevant domain of phenomena—that is, human action and thought” (p.68). There are other views of these same abilities that work to avoid the term ‘theory’ (see Harris, 1991; Johnson, 1988), but there is still broad support for the idea that children’s early grasp of certain aspects of mental life represents an organized and sophisticated area of knowledge.¹ In what follows, I will briefly review some of the specific evidence that has been reported regarding children’s understanding of false belief and discuss the theory of mind that such an understanding is said to reveal.

Evidence of Children’s Understanding of False Belief

Wimmer and Perner’s (1983) “Unexpected Transfer” task is widely regarded as the first procedure specifically designed to assess young children’s grasp of false belief. In this now classic procedure, Maxi and his mother (typically puppets or dolls) return from a shopping trip with some chocolate. The chocolate is placed in a particular cupboard and Maxi leaves the room. In his absence his mother moves the chocolate to a second location. As Maxi prepares to return to retrieve the chocolate, the subject is asked to predict where Maxi will look for the chocolate or where Maxi thinks the chocolate is located. Wimmer and Perner’s initial report pegs the age at which children can pass this task by responding that Maxi will look in the wrong place at roughly 6 years. Subsequent modifications and replications have succeeded in driving this threshold back by about 2 years (Hogrefe, Wimmer & Perner, 1986; Perner, Leekam & Wimmer, 1987; Wimmer, Gruber & Perner, 1984). Similar findings resulted from the use of related procedures such as the “unexpected contents” task in which a box of “Smarties” or other candies is actually revealed to

¹ A notable exception to this general consensus is found among those who view cognitive development from a nativist perspective. Theorists such as Fodor (1983) and Leslie (1987), for example, argue that the acquisition of an understanding of mind is an instance of modularity and that false belief understanding and its developmental precursors (e.g., social referencing in infancy and pretense at about 18 months) are “built-in” capacities that come
contain some unexpected item such as pencils. In these tasks, subjects are asked what some naive person will think is inside the now closed box. Children younger than 4 years of age report that such persons will automatically know that there are pencils in the box and even that, prior to seeing the contents for themselves, and after having voiced the reasonable belief that Smarties were to be found in the box, they nevertheless will claim that, from the outset, they themselves always believed that the box contained pencils. The conclusion regularly drawn by the authors of these studies, was that, prior to the age of 4 years, children lack the conceptual means to appreciate that persons can be led to form and act upon beliefs that are objectively false.

This conceptual deficit view was met, perhaps not unexpectedly, by a wave of counter efforts meant to champion the skills of still younger children and to promote a more liberal “early onset view” according to which 3-, and even 2-year-olds should be credited with some early or fledgling theory of mind (Chandler et al., 1989; Freeman et al., 1991; Fritz, 1992; Hala et al., 1991; Lewis & Osborne, 1990; Zaitchik, 1991). While this is clearly not the occasion to review the success or persuasiveness of these more recent techniques and arguments, it should be noted that, whatever else may divide them, all share an emphasis on false belief understanding as central to any theory-like appreciation of mental life.

What all of these reports bring out, whether emanating from the early– or late–onset camp, is that the continuing emphasis on false belief and the representational basis of children’s early insights into other minds reinforces the view that the child’s first theory of mind is based on a representational epistemology. That is, by concentrating on the child’s ability to conceive of the mind as a representational device, we ascribe to them a theory of knowledge according to which knowledge consists in the generation of accurate mental representations that correspond with increasing fidelity to a given reality—a reality that exists outside of our efforts to come to know it. This “copy theory” (Chandler & Boyes, 1982) or representational epistemology will be set against a constructivistic epistemology that is, if not fully anti-representational, then at least constitutes a

‘on-line’ at different points in the life span.
more closely embodied and interpretational view of knowing. Here reality is not pre-given, and cannot be so easily separated from the form of intelligence that seeks to know it. Rather, on this account, knowledge is understood to be constructed and jointly shaped both by the object of knowledge and the knower. This more hermeneutic view holds knowing to be dependent on notions of critical interpretation, consensus, and dialogue between persons, between persons and the environment, and between points of view or interpretations. None of this, however, should be taken to suggest that persons do not have and hold mental representations, or that representations become supplanted at some point in development by interpretations. In contrasting “representational” and “interpretive” views, I mean only to distinguish between an understanding of the separable contributions made to acts of knowing by the object and the subject—a distinction that is ordinarily taken to be definitional of what it means to ‘be’ interpretive.

These weighty philosophical debates cannot, of course, be settled here, and the studies reported in this document are not offered as a critical test of the adequacy of constructivistic epistemologies over more limited representational ones. Instead, what these studies are meant to show is that there is a period of time in development after children have acquired a representational view of knowledge—that is, after they have begun to appreciate the possibility of false belief—but before they begin to appreciate (if only tentatively) what von Glasersfeld (1988) calls one of the basic principles of constructivism: that “knowledge is not passively received either through the senses or by way of communication, but is actively built up by the cognizing subject” (p.83). Thus, attention is focused much more narrowly on a comparison of false belief understanding with an understanding of interpretation, and on the question of whether children who view the mind as a representational medium should be automatically understood to also (and automatically) appreciate that the mind is interpretive.

Two things need to be accomplished in advance of any direct comparison. The first is to show that there really are persons who equate these abilities and who take the presence of one to be sufficient evidence of the presence of the other, and the second is to show that there are principled reasons to believe that such a view is, in fact, mistaken.
Equating Interpretation with False Belief

As noted above, there are researchers who appear ready to grant an understanding of interpretation to any child capable of passing a false belief task. In characterizing the preschool years as a period in which children acquire an "interpretational or constructive understanding of representations," Wellman (1990), for example, would seem to leave little room for further development. In conceding that 2 1/2 to 3-year-olds are "novice not expert theorists," Wellman goes on to state that such expertise arrives in relatively short order. Citing studies of 4- and 5-year-olds, he ascribes to them "a notion of mind as a system of interpretive mental processes that indirectly produce not only thoughts, beliefs, and desires but hunches, conjectures, confusions, ambiguities, and subtle misconceptions as well" (p.90). Four-year-olds then are said to share with adults an understanding of knowing that embraces such concepts as "construal, interpretation, conjecture, and hypotheses" (p.89).

For his part, Flavell (1988) reads some of this same evidence—the 4-year-old's understanding of false belief—as a special case of understanding the "appearance-reality" distinction. This latter skill is said to demonstrate that children know that "other people have internal experiences that are cognitively connected to external objects and events [and] these cognitive connections engender inner, mental representations of the external objects and that the same object can be represented in different, seemingly contradictory ways" (p.247). In describing what is required to successfully negotiate both sorts of tasks, Flavell notes that, in grasping the "one-many" relation between things and representations of things, the subject must be able to "think of the other child as an active interpreter or construer of content rather than a passive receiver of content, and to understand that the same content can have two different, mutually contradictory mental representations" (p.253). More recently (Flavell, Green & Flavell, 1993; Flavell, Green & Flavell, 1995), he has adopted the view that these developments may, in fact, be more protracted, and that more fully interpretive perspectives arise in degrees rather than a single giant step, as children become more adept at detecting the "top down" mental origin (as opposed to "bottom up" environmental basis) of such differences where the procedural cues are "less salient and more subtle."
Unlike Wellman and Flavell, who now appear ready to temper earlier claims, Josef Perner has remained consistent and explicit in equating false belief understanding with an appreciation of the interpretive nature of the mind: "at around 4 years children begin to understand knowledge as representation, with all its essential characteristics. One such characteristic is interpretation" (Perner, 1991, p. 275, italics original). What remains to be accomplished, beyond the preschool years, is the ability to "combine [mental] states recursively for attribution of second-order states, for example, beliefs about beliefs, intentions about beliefs, and the like" (Perner, 1988b, p.271). For Perner, this recursive skill apparently constitutes the only developmental change of interest beyond the attainment of false belief understanding. He bases this conclusion on evidence meant to equate false belief with interpretation by showing that children understand the mind as "an active interpreter of incoming information" rather than a passive recorder of events. Such evidence is said to come from demonstrations that "a listener will use existing knowledge to actively evaluate the truth of a message" (Perner, 1991, p.53) and will discount a message that conflicts with a firmly held belief to the contrary.

It is clear that what Perner has in mind by "active" and "interpreter" is the ability of persons to evaluate the truth value of representations or to assess the relative informativeness of information in coming to know something (i.e., in forming an accurate representation of reality). In both cases, the test of this ability is found in the acceptance or rejection of a proffered belief—for example: "Will person A believe what person B says?" or "Does person X know fact Y?" Although more will be said about these matters, the essential point to be noted here is that questions of the foregoing sort address only children's understanding of the possibility of false belief—of truth and lies—rather than their ability to detect multiple interpretations or differences in the way meaning is construed.

The more general contention to be promoted here is that the contemporary literature on children's developing beliefs about beliefs is characterized by a deep running confusion about what it could reasonably mean to talk about an "interpretive" as opposed to a "non-interpretive" theory of mind. The root of this confusion, it is proposed, is to be found in a collective failure to exercise
adequate care and concern in our attempts to calculate when it is that two views can be properly said to be different views of one and the same thing. There is, for example, a certain loose, but nevertheless familiar way of speaking that gives us license to talk about the fact that Perner's standard protagonist “Maxi” and Maxi’s mother each have different beliefs about the self-same matter of where Maxi’s chocolate happens to be at certain critical moments, just as it seems appropriate to describe “one and the same” item from an appearance-reality test as having been perceived to be a rock by some, while known to be a sponge by others (Flavell, 1986; Flavell, 1988). The “loose” part in all such easy talk about identity is that while the location of the chocolate and the identity of the rock-sponge used in these examples may be in some sense unitary and fixed, the event knowledge available to the protagonists whose viewpoints are in question is, in fact, experientially quite different. Maxi’s mother, in the standard example, has access to all of the relevant facts of the matter, some of which are lost upon her son, who is strategically made to be out of the room at just the crucial moment when the question of where the chocolate is really located is finally settled. Consequently, while the two are fully entitled to their separate opinions, their different beliefs, while broadly having to do with the “same” topic, don’t actually reference “one and the same” reality. In fact, the whole thrust of this and other so-called “privileged information” or “unexpected transfer” tasks is to carefully engineer circumstances such that what is revealed to one story protagonist is pointedly different from what is accessible to some second story character. In exactly the same way, the onlooker who has only seen, but not yet touched the sponge-rock in the false belief version of Flavell’s task, is not playing with the same deck of informational cards available to those who are already in on the joke. Of course, for certain purposes, none of this makes any difference. Where it does make a lot of difference is when one is

---

2 Flavell’s procedure makes use of a sponge that has been cleverly painted to look like a rock. Children are allowed to manipulate the “sponge-rock” and are then asked about its appearance: “When you look at this with your eyes right now, does it look like a rock or does it look like a sponge?” The child is then asked the Reality question: “What is this really and truly—is it really and truly a sponge or is it really and truly a rock?” This becomes a false belief task when one asks subjects whether someone who has only seen, but not touched, the sponge-rock will think it is a sponge or a rock.
trying to establish that point in development at which children come to appreciate that one and the 
*same* event can in fact support multiple interpretations.

If all of this is correct, then individuals whose operative theories of mind are still "copy theories" (Chandler & Boyes, 1982) in which knowledge comes from objects rather than subjects, should have no special difficulty with standard false belief and appearance-reality problems, all for the reason that these measures strategically turn on the fact that each of the persons whose beliefs are in question have access to different information. Nothing about the performance of subjects on such "privileged information" tasks tells us anything, one way or another, about their capacity to: (a) appreciate that beliefs are underdetermined by matters of fact; (b) that there is a mind-to-world direction of fit operating in our epistemic lives; and (c) that two persons, both of whom have access to *precisely the same* information, can still end up holding to quite different beliefs about their common experience. Consequently, if the task at hand is to somehow decide when it is that young persons actually do come to subscribe to a theory of mind that is legitimately interpretive in this broader and more usual sense, then we will need to pursue that question using procedures other than those that currently make up the roster of standard measures of false belief understanding.

A list of required ingredients for the construction of any persuasive test of the hypothesis that an interpretive theory of mind is both different from—and later arriving than—false belief understanding would necessarily include the following:

a) a sufficiently refined knowledge about the likely course of children’s epistemic development to allow for a guided choice of study populations;

b) access to assessment tools and strategies that are generally agreed upon as adequate to the task of identifying whether given children do or do not as yet possess an understanding of the possibility of false belief, and, most importantly;

c) the availability of measurement tools capable of discriminating between young subjects who do and do not subscribe to the possibility that two individuals outfitted with identical access to precisely the same stimulus event could, nevertheless, end up interpreting this same fact of the matter differently.
The first two of these prerequisites are easily met. Although there is ongoing debate about the earliest age at which children first come to an understanding of the possibility of false belief (for a review, see Moses & Chandler, 1992), what is largely beyond any shadow of doubt is that children who are 5 years of age or older can ordinarily be counted on to pass any and all tests of false belief and so deserve being credited with possessing such an understanding (Astonington et al., 1988; Frye & Moore, 1991; Lewis & Mitchell, 1994; Perner, 1991; Wellman, 1990). The problem space within which any search for children who clearly possess false belief understanding, but do not yet subscribe to a more fully interpretive theory of mind, is consequently framed within a window involving early school-aged children who can be expected to pass standard false belief measures, but whose understanding of the interpretive character of knowing remains untested. Second, while there is still room for disagreement about what might count as a “minimally complex” measure of false belief understanding (Chandler et al., 1989; Freeman, 1994; Freeman et al., 1991; Fritz, 1992; Hala et al., 1991; Lewis, 1994; Lewis & Osborne, 1990; Mitchell & Lacohee, 1991; Moses & Flavell, 1990; Siegal & Beatty, 1990), there is no shortage of experimental tasks that are broadly acknowledged to be definitive, even if not minimally complex. Tests of preschoolers’ ability to recognize that closely held and happily acted upon beliefs can nevertheless prove to be mistaken. What is just as clearly missing, however, and what many of the energies to be expended in this study sequence are meant to remedy, is anything that could be made to count as a fair test of children’s emerging understanding that two persons with access to the same stimulus event might still come away from such a common experience with sharply different beliefs.

In seeking some counsel from the literature on this last matter, several sets of studies—both old and new—appear to hold some real promise. These are reviewed below with an eye toward the extent to which they can be seen as legitimate measures of children’s understanding of interpretation, and for the subtleties they reveal about the relations between representational and interpretive views of knowledge.
NARROWING UPON AN APPROPRIATE SET OF ASSESSMENT TOOLS

Among the numerous options that present themselves as potential procedural means of assessing children's emerging capacity to appreciate the interpretive character of the knowing process, two possibilities especially recommend themselves. One of these turns upon the fact that it is possible to identify a small class of stimuli that, whether by nature or design, turn out to have the unique feature of reliably prompting two—and only two—especially likely interpretations. Homophones are instances of this class for the definitional reason that the identically sounding words that make up this class naturally support just two distinct interpretations. When someone asks that you “take the watch,” for example, you should feel equally entitled to pick up the timepiece or otherwise stand guard over it. Certain line drawings, for example, Jastrow's “duck-rabbit” drawing (Jastrow, 1900), and Bugelski's “rat-man” (Bugelski, 1960), similarly share the property of being easily taken to be one or the other of just two different things (see Figure 5, p.35). The second general class of stimuli that provide especially fertile ground for soliciting multiple interpretations is made up of things like Rorschach inkblots and clouds and puddles of spilled milk that, because of their vague or amorphous natures, appear to easily call out a wide variety of different responses in different people. Stimulus materials of the first variety made up the working ingredients of a study sequence recently carried out by Carpendale and Chandler (Carpendale, 1995; Carpendale & Chandler, in press). In contrast, the research conducted here adopts the second of these two measurement strategies by featuring as stimuli a series of line drawings that, because of their fractional and ill-defined character, seem especially likely—at least in the eyes of those who are already interpretive—to prompt different people to regard them differently. In promoting the merits of these complimentary approaches, several alternate methodologies will be reviewed with special attention given to the extent to which each can be seen as an opportunity for children to express an understanding of the fact that exposure to precisely the same information about some stimulus event can give rise to more than one belief or interpretation. These studies are taken up in chronological order, beginning with a series of studies by John Flavell that concern children's perspective-taking abilities.
Promising Leads

**Perspective Taking Skills**

Flavell's studies of what was then called "Level 2 Perspective Taking" were chiefly concerned with children's' ability to appreciate that different perspectives (typically visual perspectives) can result in different meanings being assigned to one and the same stimulus (Flavell, 1978; Flavell, Everett, Croft & Flavell, 1981; Masangkay et al., 1974). One of the ways in which this competence was assessed consisted of presenting young children with a line drawing of a turtle that was laid out in front of the subject "feet first," so to speak (Masangkay et al., 1974). The critical question that was posed to young subjects concerned "how" that self-same turtle would seem to someone who sat opposite them and whose approach to the turtle picture was, consequently, back to front. Interestingly, 4-, but not 3-year-old subjects responded to the rather cryptic question "how does X see the turtle?" by reporting that "X" saw it "as if" lying on its back. Clearly, this is a study which, at first blush, seems to be of the sort we are after; one meant to inquire about different accounts of one and the same stimulus object. What is unique and somewhat troublesome about this particular experimental set-up, however, is that it too easily supports a confusion over two possible meanings of the key word "how." What we really want to know when we ask children how "X" sees some particular stimulus (in this case, the turtle), is whether they do or do not appreciate something like the fact that different people often find different meanings in the same objective matter of fact. Unfortunately, the turtle task would appear to conflate this question with a different and perhaps much simpler one about angles of regard—a question that could be literally gotten to the bottom of by walking around to the other person's side of the table. One could not, I take it, hope to similarly get at how you or I might see issues such as capital punishment or abortion by turning the tables in that sort of literal way. Turtles, it would seem, from this side of the table, threaten to lead us down a blind alley.

Related studies of visual perspective taking fall prey to this same objection by strategically arranging matters such that the information that is made available about the display to each observer
is distinctly different—revealing only an awareness of the effects of different visual experiences of the same object, which is not at all the same thing as recognizing that the same visual experience can be assigned different meanings.

**ROLE-TAKING SKILLS**

Within this same research tradition that concerned itself with children’s perspective-taking abilities, Chandler and Helm (Chandler & Helm, 1984) set out to add a more interpersonal dimension to the procedures then in common use by seeking to resolve the place of “shared experience” in children’s developing role-taking competence. Their study began with the premise that in assessing role-taking skills one can control the degree of overlap between the subjects’ experience of any display and that of the target ‘other’ by strategically manipulating the information that is afforded from each participants’ vantage point. The extent to which children are made sensitive to this match (or mismatch), in any given testing context can either help or hinder their efforts to report on perspectives different from their own. Chandler and Helm’s hypothesis concerned “the assumption that while having stood precisely in another’s shoes may be prerequisite to successful role-taking during certain periods of development, there are other times before which such shared experiences are helpful and after which they are no longer necessary” (p.147). Because Chandler and Helm made use of some of the same stimulus materials that were employed in the studies conducted for this dissertation, some considerable attention will be devoted to the particulars of their procedure.

The procedure made use of a set of line drawings called “droodles”—a cryptic line drawing or puzzle picture popularized by cartoonist Roger Price (1953). Such drawings are perhaps better illustrated than explained. Figure 1, for example, depicts a variation on a droodle originally published by Price over the caption “A ship arriving too late to save a drowning witch.” The humor in this and related drawings by Price is obviously meant to arise from the fact that, given the restricted or keyhole view imposed, it would be ludicrous to imagine that anyone could ever correctly guess about the larger scene of which the droodle itself is only a small, nondescript part. Once oriented by the caption, however, the cryptic fragments of the original drawing fall into place.
and it becomes possible to imagine—even difficult not to imagine—that the drawing depicts some fractional part of what its caption announces it to be.

*Figure 1: Droodle labeled “A ship arriving too late to save a drowning witch”*

The procedures employed by Chandler and Helm made use of a set of such cartoon drawings patterned after those developed by Price and extended in such a way as to include not only the limited details contained within the original droodles themselves, but also the more complete scenes suggested by the droodle captions. Each drawing was fitted with a cardboard overlay into which a small viewing window was cut. Once in place, these overlays masked most of the extended drawing, exposing only the original droodle portions of the picture, as illustrated in Figure 2.

*Figure 2: The “Ship-Witch” droodle.*

A total of six droodle stimuli were used including, in addition to the ship-witch droodle shown above, depictions of two elephants smelling a grapefruit, a camel, a rabbit, a man playing a
trombone, and an ostrich with its head in the sand. Four-, 7-, and 11-year-old subjects were tested under one of two conditions. In the first condition, subjects were shown a picture and asked to describe its contents. The cover of the picture was then closed to reveal the restricted (or droodle) view and the subject was asked “to describe how this abbreviated or abridged picture would be seen and understood by one of their classmates, who they were told would not see the completed drawings, but only the droodle portion of the picture that was then in front of them” (p. 151). According to Chandler (personal communication, 1994), the actual test question was “What will [named classmate] think this is?” Condition 2 was identical, with the exception that before seeing the full picture, the subjects were shown the restricted view and asked “What do you think this is?” (Chandler, personal communication, 1994). In each case, it was assumed that “any attribution to one’s role-taking partner which implied that such a minimally informed bystander would know, as the subject legitimately knew, the larger meaning of the unabridged droodle picture would constitute an unambiguous egocentric error, and would demonstrate a flaw in the ability of such subjects to adopt the perspective of someone less well informed than themselves” (p.151).

The subject’s responses were scored in a manner that assigned points based on the relation between the response and the true contents of the picture. Three scoring categories were used. Subjects received 2 points if they responded that their classmate would simply know the true contents of the picture. No points were awarded if the subject attributed what would now be called a false belief to the classmate, that is “a sharp distinction was maintained between the subject’s own privileged understanding of the complete picture stimulus drawing and the necessarily limited and different understanding of persons seeing only the masked or droodle portion of the picture” (p.153). Between these extremes fell what constituted contamination errors. Here subjects were given a single point if their response “inappropriately ascribed to others indirect knowledge about the stimulus drawings to which they had no legitimate access” (p.153). This category would include, for example, saying of the ship-witch droodle pictured above, that the classmate would see “two dunce caps” or “two boats on a collision course.”
Based on a set of 6 trials, subjects received scores ranging from 0 (perfect perspective-taking) to 12 (complete egocentrism). The findings with respect to the eldest and youngest groups subjects were rather straightforward: 4-year-olds proved to be essentially incapable of taking another person’s perspective by relentlessly attributing knowledge of the full picture to those who saw only the restricted droodle view, while 11-year-olds performed nearly flawlessly by giving imaginative renderings of the droodle view that strategically avoided revealing any hint of their own privileged knowledge.

The data of special interest to the present purpose come from the 7-year-old subjects. In keeping with their main hypothesis, the authors report that the performance of 7-year-olds is enhanced by having them first decide for themselves what the droodle might represent: Mean scores in Procedure 1 were 7.8 of a possible 12, as opposed to 3.5 in Procedure 2 (recall that lower scores indicate better perspective taking). Overall, however, the mean score per trial for the 7-year-olds was .94, suggesting either a substantial number of contamination errors within this group, or a stratified sample containing some subjects with scores of 2 and others with scores of zero. Chandler (personal communication, 1995) has noted that the criteria for contamination were particularly wide ranging: “shark fins” as a reading of the ship-witch droodle, for example, would be ruled a contamination error by virtue of its marine theme. Nevertheless, even if these rules were relaxed, the 7-year-olds in this study appear, as a group, to be providing fewer ‘false belief’ responses than would be expected given that even 4-year-olds routinely pass measures of false belief understanding. This suggests that when the procedure is viewed as a false belief task, it is a rather stringent test of false belief understanding. Notably, the scores for subjects in Procedure 2 are much better—a finding that the authors attribute to these subjects having first been asked about their own thoughts about the droodles. This last finding will be revisited later as attempts are made in my own studies to enhance children’s performance through various training schemes.

The useful lesson to be taken from this procedure is that droodles afford two quite different assessment opportunities. First, these stimuli allow us to assess children’s ability to appreciate that without privileged access to the full picture, a naive observer will be led to hold what amounts to a
false belief about the larger drawing. Second, the same procedure holds the potential to assess children’s understanding of interpretation since the droodles should, by design, support many different readings. The subjects in Condition 2, for example, may ‘see’ one thing in the droodle while imagining that their classmate would ‘see’ the same droodle differently. Unfortunately, Chandler and Helm’s report makes no mention of the similarities or differences between subject’s responses for self and classmate. In principle, then, this procedure would seem to have all the necessary ingredients of a test of interpretation: the subject and the classmate end up looking at precisely the same image (the droodle) and yet are free to come to entirely different conclusions as to what this collection of lines might mean. To be considered a true test of children’s interpretive capacities attention would need to be shifted from the relation between the true contents of the picture and the response given for the classmate (i.e., does it constitute a legitimate or uncontaminated false belief), to the relation between two naive readings of the same droodle (i.e., can more than one belief be had about this particular droodle?). In ways that will be detailed later, the studies carried out here attempt to more fully exploit this potential feature of the Chandler and Helm procedure.

**Conceptual Perspective Taking Skills**

Droodles reappeared in the role-taking literature in a 1988 report by Marjorie Taylor (1988), this time billed as a measure of “conceptual perspective taking” and children’s appreciation of the “mentality-reality distinction.” In a pair of studies, children viewed pictures containing, for example, an elephant and a giraffe, while puppets viewed restricted (or droodle) views of these same pictures. Subjects were then asked (when only an unidentifiable portion of the elephant could be seen in the restricted view) whether the puppet would or would not know that there was an elephant in the picture. Taylor describes two developmental levels in children’s ability to report the “interpretation” of someone who has less background information about what she terms a shared visual event:

“At Level 1, children have difficulty separating their own knowledge or interpretation of what they see from the perceptual information that is given in the environment. Children at this level would understand that if someone does not see
an object, then that person does not know about it; however, they would tend to believe that if two people both see an object or event, their knowledge or interpretation of it is identical. Level 2 conceptual perspective taking refers to an awareness that the same visual information can be interpreted in different ways. At Level 2, children can take differences in background knowledge into account when determining how people will interpret a perceptual display.” (p.704)

Taylor’s use of the term “interpretation” suggests a somewhat different sense than the one I have been employing. In the quotation above, as well as elsewhere in the report, she appears to equate the terms ‘knowledge’ and ‘interpretation.’ Because her purpose is to chart the development of children’s understanding that someone lacking relevant background information cannot know the identity of objects contained within a droodle view, Taylor’s studies do not directly address the key difference between knowledge and interpretation, namely, that persons with access to precisely the same information can still come to hold different beliefs about that same reality. To understand that someone will not know that there is an elephant in the picture is not the same thing as understanding that the puppet is free to ‘see’ one thing in the droodle, while you or I (if denied access to the full picture), are free to ‘see’ another. Level 2 understanding hinges on an appreciation of differences in background knowledge rather than on the characteristics of the stimulus picture that afford multiple interpretations or the inherently interpretive nature of the knowing process itself. Still, Taylor’s distinction between Level 1 and Level 2 understanding comes closest to capturing the essential ingredients of an interpretive theory of mind and, in extending the findings of Chandler and Helm, she reports both training effects and limits on the kinds of information that children are willing attribute to the uninformed puppets. For these reasons, a more detailed description of the studies is offered below.

Taylor frames her studies as a test of what she calls the “seeing = knowing” hypothesis, that “young children make no distinction between experts who know exactly how to examine a display and appreciate the significance of what they see, and novices who have no special scanning strategies and do not know what different parts of the display signify” (p.703). As a follow-up to the Chandler and Helm study, Taylor set out to determine whether, if young children really do equate seeing with knowing, they would, as a consequence, attribute knowledge of the full picture to a puppet who sees a restricted view that reveals only the tiniest of edge of the object. In
extending these previous findings, she also asks whether children will believe that droodles are sufficient to impart other kinds of knowledge such as the actions being performed within the picture (the elephant is sitting down) or personal information about the animals depicted (the elephant’s name is George). Finally, Taylor conducted a second study that included a set of training trials meant to improve the performance of 4- and 6-year-olds.

An example of the stimulus pictures used in Taylor’s first study is shown in Figure 3 below. The 5 restricted views of the same picture are shown on the right-hand side of Figure 3. The “Empty” restricted view was used to determine whether or not children would think that even an empty window is sufficient to allow identification of the objects contained in the larger picture. The restricted view labeled “Tiny Edge” was meant to assess “how far it was possible to push the seeing = knowing hypothesis... [that is, would]... children attribute knowledge of the object to an observer who sees only a tiny edge” (p.707). Using the “One Object” view children could claim that the puppet can identify the object that is partially in view but not another object in the picture that is completely occluded. The “Two Objects” and “Identifiable Part” views were included to distinguish subjects who might have a bias toward reporting either that the puppet always knew about only one of the objects (Two Objects), or was always in complete ignorance about the picture (Identifiable Part).

The procedure for Study 1 involved showing 3-, 4-, 5-, 6-, and 8-year-old subjects the full picture, then showing one of the restricted views to the puppet and asking “Does [puppet’s name] know there is a [giraffe] in the picture?” On two of the 6 trials, the children were asked about the puppet’s ability to know other pieces of information to which the child—but not the puppet—had prior access: “Does [puppet’s name] know the giraffe is sitting down?” and “Does [puppet’s name] know that the giraffe’s name is George?” The results of these inquiries were scored to produce a percentage of subjects in each age group who consistently gave “Level 2 responses”—those subjects who responded that the puppet would not know there was a giraffe in any picture save the one in which an identifiable part of the giraffe is visible, and that the puppet would not know the giraffe’s name or the actions visible only in the full picture. If one recalculates Taylor’s data as
reported, and allows each subject one mistake (that is, one Level 1 response) in the series of 6 trials, the results appear as shown in Table 1.

Figure 3:  *Stimuli used by Taylor (1988, Study 1).*

![Diagram of stimuli](image)

**Table 1:**  *Percentage of subjects giving one or fewer Level 1 responses (adapted from Taylor, 1988).*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Percentage of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-year-olds</td>
<td>50.0</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>60.0</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>62.2</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>68.7</td>
</tr>
<tr>
<td>8-year-olds</td>
<td>93.7</td>
</tr>
</tbody>
</table>

These results can be read as a partial index of the subjects' capacity to know that the puppet will have a false belief as to the contents of the picture. This follows if only for the reason that a prerequisite to attributing a false belief is knowing that the puppet will not know the identity of the object. One might wonder then, why these subjects—especially 5- and 6-year-olds—appear to have had such difficulty appreciating that the puppet cannot know the identity of the object given their restricted view. This is particularly puzzling since 4-year-olds, in general, are known to do
well on other measures of false-belief understanding, and these particular children appear to have been well on the road to false-belief understanding: all subjects were pretested to ensure that they could appreciate what was and was not visible in the restricted views; those 5 years of age and older “almost never” reported that the puppet could identify the object given the “tiny edge” view and they rarely attributed action knowledge or personal knowledge to the puppet. Why, then, did a substantial number of these subjects fail to act as though they appreciate the possibility of false belief, but instead claim that the puppet would know the identity of the object given only a key-hole view? A possible explanation may lie in the fact that, although Taylor notes that having highly distinctive features of the objects visible through the window may lead children to (wrongly) believe that a puppet could identify the object, she nevertheless chose to reveal features that were rated by a sample of adults to be “very or absolutely distinctive for a given object” (p.705). The lesson taken from this finding and applied in my own studies, is that droodles should reveal features that are less distinguishing or definitional.

Taylor, it seems, was also puzzled by these results and went on, in Study 2, to test the hypothesis that part of the reason for this unexpectedly high failure rate is that children of this age “may not think to search for alternative perspectives to their own or consider the possible interpretations that an ambiguous surface structure might engender” (p.712). The purpose of the study was to determine if children would benefit from a demonstration of how “the possibility of multiple interpretations of a stimulus is relevant in the context of perspective taking” (p.712).

The study was comprised of two conditions. In both conditions, 4- and 6-year-olds were given a series of restricted-view tasks similar to those used in Study 1, but subjects in the Training Condition were “alerted to the possibility that restricted views were potentially ambiguous” (p.712). This was accomplished using a series of different pictures that could be masked to produce identical restricted views (see Figure 4, below). For example, the restricted view of a picture of a boot showing just the toe of the boot was identical to the restricted view of a sinking ship showing just a smokestack, and to the restricted view of a pig showing only the snout. Those in the Training Condition watched as a puppet failed to identify the real objects from these identical
restricted views—mistaking the boot for the pig, etc. Training ended when the child responded on three consecutive trials that the puppet would not know which of the three objects was behind the restricted view (or after 9 trials). Those in the Control Condition watched as the puppet failed when seeing non-identical restricted views of these same stimulus pictures.

*Figure 4: Restricted views used by Taylor (1988, Study 2).*

Children in both conditions were then shown a series of 6 pictures (all with different restricted views) and asked, for each one, “Does [puppet’s name] know this is a _____?” The mean number of correct responses (out of 6) to this question appears in Table 2. Taylor reports a statistically reliable difference between the conditions and concludes that this “shows that children seem to learn that restricted views in general can be ambiguous” (p.716).
Table 2: Mean number of correct responses by Age Group and Condition (adapted from Taylor, 1988).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>3.5</td>
<td>1.8</td>
</tr>
<tr>
<td>4-year-olds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-year-olds</td>
<td></td>
<td>5.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

While the term “interpretation” is used throughout Taylor’s report, in neither of the studies are children asked what the puppet actually thinks—only whether or not the puppet will know this is a picture of a giraffe, or a boot, etc.. Children’s understanding of interpretation—that persons (or puppets) might interpret the restricted views differently—is never directly addressed. Instead, the studies explore children’s appreciation that the absence of sufficient background information makes it impossible to identify the objects in a restricted view. While this can easily be seen as a proxy measure of their appreciation of at least some aspect of the concept of false belief, there are more explicit indications that some of her subjects actually did think that the puppet would entertain a false belief: “Over half of the 6-year-old children spontaneously speculated on alternative hypotheses for the identities of the parts in the restricted views. For example, children observed that someone might think the rabbit was a boomerang, submarine, airplane, or plant. Only three of the [16] 4-year-old children made this type of comment, but they were all in the training condition” (p.717). There are good reasons to suppose that had the subjects been asked to attribute beliefs to the puppet (as in the Chandler and Helm procedure), some non-trivial proportion would have given bona fide false beliefs. My own contention is that had Taylor then gone on to ask about the thoughts of a second puppet, only her 8-year-old subjects would evidence a working grasp of interpretation by giving a second, different false belief. This claim is put to the test in my own Study 3.

Although Taylor’s training trials had the stated goal of a demonstrating “the possibility of multiple interpretations” the actual training trials primarily emphasized “the possibility that restricted views were potentially ambiguous” with regard to which of the three training pictures might lie beneath. While it is true that such ambiguity might promote multiple interpretations, her
training trials seem especially aimed at influencing subject’s understanding of false belief—that a single droodle can indicate more than one picture with equal probability, rather than the more fully constructivistic notion that the droodle can support multiple beliefs that have nothing to do with the underlying picture. It is not clear, therefore, whether the training trials work to promote only an understanding of false belief, or also have the effect of highlighting the rich interpretive possibilities inherent in the stimuli. Providing empirical evidence on this matter was one goal of the present study sequence.

Knowing more about the way children themselves think about the process of attributing beliefs to others in situations that are particularly open to multiple interpretations is also a goal of these studies. Adults understand that differences in interpretation are not always the result of simply being kept in ignorance, as is the case with target characters who are never shown the full picture before the cover falls in place, but more commonly arise for ‘deeper’ reasons. Higgins (1981) has pointed out that one can distinguish cases of interpretive diversity that are due to “situational” factors and that arise from differences in recent perceptual history (what one did or did not see) and those that can be said to stem from “individual” factors such as personality traits, abilities, specific knowledge, etc., that are more enduring characteristics of persons. The droodles procedure, as implemented by Taylor and by Chandler and Helm, represents only the first of these types.

In a follow-up study focused on the second, and more ‘individual’ of these sets of factors, Taylor (Taylor et al., 1991) set out to determine if young children were sensitive to the differences in general knowledge that exist between babies, children and adults and, if so, whether such sensitivities could allow them to detect or predict that individuals of different ages will differ in their interpretation of a droodle. In a series of 3 experiments, Taylor et al. demonstrated that children are sensitive to the presence of differences in general knowledge by 4 years of age: they give evidence of appreciating, for example, that children know—but babies do not—what a teeter-totter looks like. Nevertheless, 4-year-olds (but not 6-year-olds) tend to believe that a naive observer will be able to identify an object from a restricted (droodle) view, even if that observer does not know what the object looks like (i.e., lacks what we would consider adequate general
knowledge). My own Study 6 provides an opportunity for subjects to explain, in their own words, why it is that persons might differently interpret restricted views, and, if they are capable, to comment on the role of both situational and individual factors.

THE MIND AS AN ACTIVE INFORMATION PROCESSOR

Fourth on this list of studies that hold out some promise of informing us about genuinely interpretive theories of mind, is a study by Perner and Davies (1991) that set out to determine whether young children believe that the mind is an “active” or “passive” processor of information. In taking issue with the notion that children first acquire a “copy theory of mind” (Chandler & Boyes, 1982), Perner and Davies aimed to show that “children understand that belief formation is not a passive copying of the most recent information received but involves active evaluation of that information in view of existing beliefs” (p.56). Using both verbal messages and droodles as stimuli the authors claim to show that 4- and 5-year-old children already appreciate “the need for interpretation” that such materials present and take issue with the idea that there is any light at all to be found between representational and constructivistic theories of mind.

Perner and Davies begin by noting that it is not until about 4 years of age that children first appreciate that what a person knows or does not know is determined by the availability of information and that particular senses provide particular kinds of information. What is said to be controversial, is “whether at this age children conceive of the mind as an active processor of this information or whether they conceive of the mind as a passive receptacle (Flavell, 1988), repository (Wellman, 1988) or passive recorder (Chandler & Boyes, 1982)” (p.52). According to Perner and Davies, these latter possibilities could be discounted if children could be shown to appreciate that others will set aside what they take to be bad advice or misleading information.

In their first experiment, 3- and 4-year-olds were told two stories in which a speaker informed a listener about the location of an object or the identity of an object in a box. In one story the speaker provides correct information and, in the other, misleading information. When the listener has no firm belief about the location or identity of the object, he or she should simply believe what they are told. As expected, when asked about the listener’s beliefs (“Will Mary think that’s true?”)
the majority of subjects (62% of them) responded that the message would be believed. In a second and more critical condition, the listener was said to have a firm prior belief about the objects' location or identity that was at odds with the speaker’s message. Under these circumstances, the majority of subjects (56%) held that the listener would reject the message.

Perner and Davies’ second study made use of droodles as well as John Flavell’s sponge-rock. Four-year-old subjects were shown a picture (for example, an elephant smelling an orange) and then, with the cover closed to reveal only a rectangle and a circle (the tip of the elephant’s trunk and the orange), were asked if another child who had not seen the full picture would “know that this (points to circle) is an orange?” In the sponge-rock task, children were allowed to manipulate the deceptive object and then asked if a person who only saw but did not touch the object would know its true identity (“Does Mary know what’s really in the box?”). In responding to these simple yes-no questions, nearly 80% of subjects (25 of 32) felt that such naive observers would not know the true state of affairs.

To rule out the possibility that subjects in Study 2 were somehow pressed into responding “No” to a test question that strongly emphasized the negative “has not ever seen this before,” Perner and Davies repeated the procedure but included a condition in which the target person had full access to the picture or the deceptive object. Under these conditions, 61% of the 5-year-olds tested responded that the target would know what the object was when given access to the full picture or allowed to touch the deceptive object, and would not know such truths when perceptual access was withheld.

Perner and Davies conclude that the data from their first study demonstrate that 4-year-olds understand the mind “actively” in that they already appreciate that others will evaluate incoming information in light of their current beliefs, and so appreciate that individuals with different information at their disposal will evaluate incoming information differently. It seems unlikely that anyone would seriously doubt that children of 4 already see their own or others’ minds as “active” in the broad sense of somehow generally operating upon or making timely use of the new information made available to them. While we should not be especially surprised to learn that 4-
year-olds know that Mary will reject a misleading description of the object when she has good reasons to believe otherwise, it does not necessarily follow, as Perner and Davies claim, that such children have a full “understanding of the effect of this description on Mary’s mind” (p.55). This finding seems more consistent with the notion that 4-year-olds understand that droodles and deceptive objects can engender false beliefs only among those who have not yet seen the whole elephant or hefted the sponge-rock—that is, that such under-informed persons can get things wrong.

Similarly, Perner and Davies’ studies 2 and 3 fail to tell us anything decisive about the question of when children come to recognize that minds are actually interpretive in the epistemic sense of differently imposing a personalized world-to-mind direction of fit upon experience. In both cases, children are asked whether a naive observe will or will not know the truth of the matter—“Does Mary know that this is an orange?”—instead of the more appropriate question: “What will Mary think this is?” What is at real issue, but unfortunately left unilluminated by the Perner and Davies study is the question of when children first come to realize that minds not only somehow “crunch,” in some procedural way, the evidence already at their disposal, but also go beyond this to actually demonstrate an understanding that minds do more than sum old and new information, but actively reach out instead to assign different meanings to one and the same thing. Because the children in Perner and Davies’ procedures are never directly asked about such matters, these data, rather than supporting what the authors call a “nascent interpretive theory of mind,” seem instead to fit more comfortably beneath the banner of straightforward false belief understanding.

**Biased Beliefs**

Finally, in this listing of already available suggestions about the probable developmental course that leads children to an interpretive view of knowing, there are a small number of studies concerned with the power of bias. Adults appreciate that when matters are in dispute—that is, when differing views of some singular thing are offered—the basis of such disagreements is often to be found beyond the bounds of the immediate situation. As a consequence, differences of opinion, or belief, or interpretation can sometimes be appropriately understood in terms of such
concepts as bias or prejudice—some pre-existing condition that systematically distorts the way in which interpretations are formed. For young children, one potential avenue toward an interpretive stance might be found along a path paved by those situations in which there are good reasons to suppose that persons might be primed by previous experience or aspects of personality to “see what they want to see and hear what they want to hear” well in advance of any information or evidence that might be placed at their disposal. Bradford Pillow has examined this possibility in several studies of children’s understanding of what he calls biased interpretations (Pillow, 1991; 1993; Pillow & Henrichon, 1996). Pillow’s procedures typically unfold in two phases. In the first phase, children are provided with background information about a story character that is meant to establish a bias of some sort. The character may be said, for instance, to have a strong dislike for some other character. In the second phase, a vignette is presented that contains an “ambiguous action” or event—an action that, if biases are seen to be at work, could be seen in one of two different ways. For example, a character seen holding a rabbit in front of an open cage may be presumed to be in the mischievous act of taking the rabbit out of the cage, or alternatively, if he is said to be “liked,” as helpfully putting it back into the cage. What these studies have regularly demonstrated is that 6-, and 7-year-olds, but not younger children, seem to appreciate that people’s likes and dislikes will dictate how they end up viewing a range of morally and factually ambiguous events.

Such findings suggest that, whatever else they might already know about beliefs and desires, young children’s first inklings of the idiosyncrasies of other people’s mental lives may arise during the 5- to 7-year period currently under discussion. While clearly pointing in what is seen here to be the right direction, for reasons to be more fully explored below, Pillow’s procedures always involve circumstances in which there is no real fact of the matter to be decided, that is, we don’t ever know for certain if the pictured story character is really removing or replacing the rabbit, and so a binary choice is always presented in which there is a ‘right’ way and a ‘wrong’ way of viewing the behaviour of the story character. As such, these otherwise useful studies are not
fashioned to draw out the distinction between interpretation and false belief understanding that is the focus of the studies presented here.

**Matters of Taste, Ambiguity, and Interpretation**

In view of all of the shortcomings seen to operate within these previous attempts to explore children’s first thoughts about the interpretive nature of the knowing process, some altered set of procedures is clearly needed. What is required, if we are to demonstrate that false belief understanding alone is not enough, is some procedure that would allow a within-subjects comparison. That is, a procedure that would simultaneously allow for the assessment of children’s understanding of the possibility of false belief as well as some means of assessing children’s understanding of the “one-many” relation between things in the world and their possible interpretation (Flavell, 1988). As noted above, two sorts of procedures would seem to fit this bill. The first, developed by Carpendale and Chandler (Carpendale, 1995; Carpendale & Chandler, in press), employs stimuli that readily support two (and only two) mutually exclusive interpretations. The second, the approach taken here, features stimuli that seem to ‘automatically’ elicit many possible readings. Before listing out the merits of my own procedure, the methods and findings of Carpendale and Chandler’s work first need to be examined more closely.

Carpendale and Chandler reasoned that children may find the task of expressing an early appreciation for interpretation easiest to negotiate in situations that offer strong evidence in support of each of two meanings to approximately the same degree. In other words, when the number of available and warrantable interpretations is reduced to just two, the background job of imagining alternative readings is simplified. In testing these predictions, Carpendale and Chandler presented 5- to 8-year-old subjects with false belief tasks as well as tasks involving matters of taste, lexical ambiguity, ambiguous referential communication, and ambiguous figures.

The false belief task and the problems involving matters of taste were included in an effort to explicitly show what an understanding of interpretation is not. The false belief task was meant to demonstrate that an understanding of false belief does not guarantee success on tasks that require interpretation. The task used was modeled after Wimmer and Perner’s (1983) “Unexpected
Change” task in which an object is moved from one location to another and subjects are asked to predict where a story protagonist who did not witness the change in location will look for the object. Matters of taste were included because, on the surface, differences of opinion about whether, for example, a particular food tastes good or bad, could well be seen to be differences in interpretation of what is one and the same thing. Alternatively, differences in taste can be distinguished from matters of interpretation in that the former is typically an affective evaluation of, or reaction to, an object, whereas the latter concerns beliefs about some more or less “objective” fact of the matter (i.e., what an object is or means). In addition, interpretations are typically warranted or justified by appeal to evidence and reason, whereas differences in taste are typically located within persons rather than objects and are, therefore, less likely to be challenged or to require justification. Subjects’ understanding of matters of taste was assessed by pressing them to comment on vignettes in which one character was said to think a particular kind of soup tasted good, while a second character thought it tasted bad (or that a picture was “nice” vs. “ugly”).

The tasks meant to tap subjects’ interpretive understanding involved ambiguous communications or pictorial stimuli in which story characters were said to hold contrasting beliefs about the two easily available meanings of the message or picture. Problems of lexical ambiguity, for example, involved a game in which the characters are told to “wait for a pair/pear” and one is presented as waiting to eat a pear, while the other waits for a pair of shoes. Similarly, children’s understanding of ambiguous referential communication was assessed using a hiding game devised by Sodian (1990). In the game a penny is hidden under one of three cards. A picture of a block appeared on each card: the first card contains a large red block, the second a small red block, and the third, a large blue block. Two puppets are then told that “the penny is under the card with the big block” and one is said to believe it is under the card with the large red block while the other believes it to be under the card with the large blue block. Finally, subjects were shown two ambiguous figures (the “duck-rabbit” and the “rat-man”, see Figure 5 below) and the puppets were said to endorse opposite readings of these pictures (e.g., one believes it is a rabbit, the other that it is a duck).
In all of these cases, the subjects were asked a series of questions meant to probe their understanding of the interpretive problems presented by the stimuli. For example, they were asked to explain why it was that the puppets had come to their different conclusions: “Is it OK for Mary to say , and for Maxi to say ?” and later “Why is it OK...” Following this, the subjects were told about a third puppet, Josef, who held a “deviant” interpretation about the picture or message. For example, when told to “wait for a pair/pear,” Josef claims to be waiting for an apple.

What Carpendale and Chandler found using these procedures, was that all of their 5- to 8-year-old subjects passed the false belief task and procedures having to do with matters of taste, but only the 7- and 8-year-olds passed the interpretation tasks. That is, all of the subjects could readily predict and explain why it was that Maxi would not know where the object was in the false belief task, and why it was that he would look in the wrong place. By contrast, essentially none of the younger subjects could explain why the puppets held different beliefs on the interpretation tasks and simply denied that such differences were possible or legitimate.

These findings support the hypothesis that false belief understanding is an earlier developing ability that is importantly different from the ability to understand or explain differing interpretations of an ambiguous message or picture. There are reasons to assume, however, that not everyone would be persuaded that these data demonstrate a clear developmental difference between the acquisition of an understanding that persons who are differently informed may hold different beliefs (false belief), and an understanding of the notion that the same thing can be assigned different meanings by different people (interpretation). First, the insight that we want to observe in
young children—that knowledge is interpretive—is an insight about knowing, and not about things themselves: that part of the nature of knowing is that people impose meaning. In the case of the ambiguous figures, however, all the variability in responses is laid on the features of the thing itself rather than the person who imbues it with some particular meaning. Because such stimuli so easily afford but two and only two ‘permissible’ interpretations, there is the very real danger that the object somehow thrusts interpretations on an otherwise unsuspecting subject. That is, visual illusions such as the duck-rabbit, may force their two interpretations on the subject who need do little to discover them, and rather than being made to actively construct a meaning, the subject is alternately presented with first one meaning and then the other. That is, in this special class of stimuli, representational diversity more a feature of the ambiguous message than of the hapless listener. Since these features work to move the locus of the interpretive problem away from subjects and toward objects, all of this should work to scaffold young children’s attempts to deal with the interpretive possibilities inherent in such stimuli. That this leg up is apparently not sufficient to close the performance gap between the false belief and interpretive tasks could be read by potential critics as merely indicating that it is children’s ability to read symbols that is slow to develop and not their understanding that persons constructively interpret the world around them.

Support for this alternative reading of Carpendale and Chandler’s tasks is also found in the fact that the interpretations taken by the puppets are assigned by the experimenter rather than imaginatively attributed by the child subject. What this leaves partially obscured is the important question of whether the subjects themselves appreciate the ambiguity of the stimuli, or, more simply, appreciate that Maxi could be victimized by this reading, while Mary is victimized by that. In other words, do the subjects believe that Maxi and Mary are wrestling with the inherent ambiguity of the stimulus or simply clinging to their own first impressions? Carpendale and Chandler want us to believe it is the former and that children appreciate the nature of the quandary that Maxi and Mary find themselves in, but the actual procedure gives us no compelling reason to accept this view. What seems needed, then, is a procedure that assesses interpretation by allowing the subjects to directly attribute beliefs to other persons, rather than having such beliefs provided
by the experimenter, and that focuses attention on the person-centered process of coming to hold particular beliefs rather than on some object-centered ambiguity of the stimulus situation. The proposed methodology to be detailed in the next section is meant to incorporate these improvements on Carpendale and Chandler’s techniques.
A PROPOSED METHODOLOGY

To assess young children's grasp of the principle that two or more persons might differently interpret one and the same stimulus event, a set of 6 studies were conducted that employed variations on the droodles procedures developed in previous studies of children's role- and perspective-taking skills (Chandler & Helm, 1984; Perner & Davies, 1991; Taylor, 1988; Taylor et al., 1991). A review of these previous reports suggests that the procedure can be put to use as a measure of false belief understanding, that is, as a test of whether children know that viewing only a small portion of a larger picture can lead someone to hold a mistaken belief as to the picture's likely contents. In the study sequence to be reported below, a variation of this same procedure is used to assess children's understanding of interpretation by asking them what each of two naive observers will think of the same restricted view. Children's reactions to these inherently ambiguous materials hold the potential to clearly distinguish between a theory of mind that rests entirely on notions of false belief (i.e., one that views the mind as a recording device and that makes room only for mistakes and not for misinterpretations), and a more complex appreciation of the mind's capacity for interpreting and constructing (and so for misinterpreting) reality. What these studies were hypothesized to reveal is that an interpretive theory of mind makes its first appearance only early in the school years—well after a point in development when children already appreciate the possibility of false belief.

The procedures employed in all of the studies described below made use of a set of such cartoon drawings patterned after those developed by Price (1953) and extended in the same manner used by Chandler and Helm (1984). The new use made of this general procedure was to ask children how, not just one but two, onlookers or protagonists might construe the same restricted, or "droodle" view. Inquiring into the thoughts of a second onlooker succeeds where earlier studies have failed in offering subjects the chance to showcase a more sophisticated grasp of mental life under which both protagonists are potentially free to entertain different false beliefs or different interpretations of the self-same reality.
Each of the 6 studies, to be discussed in turn, contain variations on this same theme of asking what first one puppet, and then the other, would think of the same picture. The key hypotheses under examination were that false belief understanding is: (1) an earlier occurring; and (2) a necessary but not sufficient condition for the emergence of an interpretive or constructivistic theory of mind.

The study sequence began by testing a number of assumptions about how adults react to the stimuli and then proceeded by examining the various ways in which young children can be afforded opportunities to reveal their best insights into the interpretive nature of the knowing process. A sample of adults was first used because what is obviously required of any such set of ill-defined, and so, ambiguous stimuli is that, when offered up to groups of adult, or other presumably “interpretive” subjects, they should actually lead such target persons to interpret them in a wide variety of different ways. The experience of Herman Rorschach, an obvious progenitor of all subsequent attempts to intentionally fabricate ambiguous stimuli open to multiple interpretation, teaches us that the success of any such tool building effort is far from automatically guaranteed. Notwithstanding the craftsmanship with which Rorschach’s stimuli were originally produced, experience has shown that certain of these ink blots do and others do not regularly call out what have come to be described as “popular” responses—responses that are so commonplace as to be almost everyone’s first interpretive effort. What such evidence brings out is that, good intentions aside, the job of producing stimuli that are especially open to multiple interpretation is a

3 According to Klopfer and Kelly (1942), the first recorded use of ink blots for psychological purposes can be found in Justinus Kerner’s 1857 text entitled *Kleksographien*. Klopfer and Kelly describe Kerner’s “accidental” discovery of “the possibilities inherent in the use of ink blots by noting how, as he observed them, they assumed various forms, which impressed him with their bizarre meanings... Kerner’s most interesting observation was that it seemed to be impossible to produce ink blots according to a preconceived plan; rather, the ink blots tended to impose their meaning and significance on the producer. Thus he experienced the strong interplay between the objective features of the ink-blot material and the individual projections of the observer which is still the most impressive experience for everyone in a first encounter with Rorschach material. Kerner seems not to have realized fully that this interplay produces different results with different people; nor did he recognize the significance of these differences as affording the basis for a method of personality diagnosis” (p.1). Finding a set of materials meant to take full advantage of these properties would await the work of Herman Rorschach. His *Psychodiagnostik*, published in 1921 is described by Klopfer and Kelly as “the result of his ten years of experimentation [and] the selection, from among thousands of trial blots, of a standard series of ten ink-blot
task in need of careful empirical quality control. For these reasons, the stimuli to be used in subsequent studies of children were retained only where it proved rare for adult respondents to interpret them in exactly the same way.

Assuming the availability of a set of stimulus drawings that satisfy the condition of regularly prompting different interpretations in different adult subjects, it would seem possible—and perhaps even intuitively compelling—to then proceed by simply determining what it is that young subjects of various ages will suppose that first one and then another protagonist might say when exposed to such inherently ambiguous materials. If this were done, and if subjects of a certain age spontaneously reported that they (or some protagonist) would see such drawings in one way, while some second onlooker spontaneously read these same stimuli differently, then the simple fact of their having responded in this divergent fashion could reasonably be seen to count as evidence in support of their ability to appreciate that two persons can and do differently interpret one and the same thing.

Despite a certain compelling simplicity, there is, nevertheless, a subtle fault in the preceding logic that needs to be brought out—a fault that might easily result in serious confusion. The potential difficulty had in mind here turns upon the rather complicated relation that necessarily obtains between what, in any given situation, is required for a response to qualify as “interpretive,” and what is required for it to be counted as “correct.” The difficulty is this: In diagnosing whether a given individual might still regard any and all beliefs as mere copies or causal consequences of direct encounters with a supposedly “objective” reality, or, alternatively, whether that same individual views the process of knowledge acquisition as the result of some more actively “interpretive” undertaking, one might easily—but mistakenly—suppose that it is a matter of indifference whether any of the beliefs in question actually happen to be approximately true or false. That is, the job of working out whether a given target subject: (a) still subscribes to the simple and singular view that the knowing process is exclusively “mind-to-world” in its direction pictures to serve as the stimulus material in his diagnostic procedure.” (p.3)
of fit, or, alternatively; (b)whether such acts of knowing have already come to be regarded as the interactive consequence of some joint process that also fits the world to the structure of the mind, might well appear to be independent of the question of whether any one of these beliefs happens to be approximately correct.

The critical methodological point that needs to be brought out here is that, however apparently orthogonal the questions of whether a belief is understood to be an interpretation and whether it is understood to be correct might appear to be in principle, they end up proving to be naturally tied in any concrete assessment effort. This measurement problem arises as a consequence of the fact that both of these contrastive epistemic stances each allow for the possibility of somehow getting things wrong. Simple “copy theories” (Chandler & Boyes, 1982), which are argued here to require no more than an awareness of the possibility of false belief, make room for this prospect by assuming that those with unhampered access to all the facts naturally get things right, whereas those who are somehow shut out from some of these relevant details automatically fall into ignorance and so end up subscribing to false beliefs. Alternatively, more “interpretive” theorists of mind, while not at liberty to talk so simplisticly about truth and falsity, do obviously make homologous provisions for marking some interpretations as well grounded and defensible, while others are viewed as unfounded and as amounting to misinterpretations. As such, young persons both with and without an interpretive understanding of the knowing process could be expected to offer up different renderings of one and the same stimulus event whenever one or the other of two protagonists in question can be said to be simply “right” while the other is simply held to be “mistaken.”

Consequently, the diagnostic task of determining whether a particular child has already come to some interpretive theory of mind cannot be carried out in ways that are independent of the validity that attaches to the particular beliefs with reference to which such a determination is to be made. This follows for the reason that although any putative demonstration that a given subject already holds to a genuinely interpretive theory of mind necessarily requires the ability to attribute to two different persons two different beliefs about a given stimulus event, such demonstrations are also sufficient if and only if both of the beliefs in question are accorded more or less the same credence.
What is pointedly excluded by this restriction are all those cases in which one of the two protagonists in question is said to hold to a legitimate or true belief, while the other is simply deemed to be dead wrong. That is, those who unarguably hold to some genuinely interpretive theory of mind might very well respond in such a fashion for the reason that under certain circumstances it is entirely reasonable to conclude that one of the protagonists in question is simply doing a good and the other a bad job of interpretation. There is, however, nothing about such a performance to distinguish it from the response pattern of other, less competent, subjects who understand no more than the possibility of false beliefs.

Here, then, is the real upshot of this complicated relation between the prospects that a given subject’s views are interpretive and that they are true. First, for subjects who are already capable of ascribing to different protagonists different interpretations of one and the same stimuli, it is a matter of complete theoretical indifference whether either or both of these interpretations are judged to be right or wrong. After all, the notion of interpretation naturally assumes the possibility of misinterpretation. Second, nothing like the same liberties can be allowed in the cases of subjects who were somehow suspected in advance to still lack the capacity to view the knowing process interpretively. Knowing that such a subject attributes a true belief to one protagonist and a false belief to another is, after all, precisely what is meant by the ability to entertain the possibility of false beliefs and, in and of itself, tells us nothing about that individual’s competence to consider the prospect that two different persons may differently, and with equal legitimacy, interpret or misinterpret one and the same thing.

What is seen to follow from all of this close scrutiny of the relations between the truth and falsity claimed for each of two different renderings of a given stimulus event, and the potential diagnostic utility of such responses is that while any data indicating that each of two protagonists are equally right or equally wrong in their understanding of one and the same stimulus may count as evidence in favour of an interpretive theory of mind, any claim that one such protagonist is right in thinking “x”, while another is wrong in thinking “y” does not. While careful attention was paid
to the credence that subjects assigned to attributions of belief in each of the studies conducted, the implications of the relations between truth and falsity are most directly tested in Study 2.

Finally, in coming to a concluding view about what ought to count as a legitimate measurement context for assessing children's emerging understanding of the interpretative nature of knowing, it is necessary to consider whatever differences still divide those circumstances in which two protagonists might be said to hold different views both of which are seen, on the one hand, to be equally right, and on the other, to be equally wrong. From a purely theoretical vantage point there is no real light between these two scenarios and either would appear to fit the diagnostic bill. More practically, however, a great deal separates these two possibilities since there simply is no evident way of structuring an assessment task in such a fashion as to make it both reasonable and likely that a child subject will spontaneously propose that each of two protagonists will hold to different but equally meritorious and veridical beliefs about one and the same thing. By contrast, it is a relatively straightforward matter to arrange things in such a way that, whatever the particular beliefs a subject ascribes to protagonists A and B, both will be understood to be equivalently mistaken. In other words, it is procedurally easier to engineer events such that two persons can be expected to be differently wrong about the same reality than it is to produce circumstances which support and elicit multiple correct interpretations. The sequence of studies reported in the next section attempt to address these issues and to build upon the earlier research efforts reviewed above.
THE EXPERIMENTS

A total of 6 studies were conducted, each of which was meant to highlight one aspect or another of an Interpretive Theory of Mind. In brief, Study 1 establishes the overall suitability of the drooodle stimuli and general experimental procedures using a sample of adults. Study 2 demonstrates the necessity of using both restricted and unrestricted views of these stimuli as opposed to generally ambiguous stimuli—that is, of allowing subjects access to the full truth about the pictures before asking them to attribute beliefs to less well-informed others. Study 3 replicates and extends the findings of Taylor (1988) in an effort to show that it is not sufficient to ask children if a naive observer (who sees only the restricted view) will or will not know the true contents of the picture, since, at best, such inquiries can only reveal the subject’s understanding of false belief. Study 3 also explores the possibility that modeling or training can improve subject’s performance. Study 4 represents the main study in the sequence and demonstrates that in a simultaneous assessment of false belief and interpretive understanding, false belief understanding appears to be insufficient to allow for the exercise of a more interpretive theory of mind. Studies 5 and 6 are intended to rule out certain reductive readings of the results of Studies 3 and 4 by addressing issues of creativity, imagination, and subjects’ access to alternative interpretations.

In Study 1, the overall suitability of the stimuli for use with child subjects is assessed by determining the degree to which adults (who are presumably already “interpretive”) actually interpret these particular drawings differently. That is, the open question that is taken up is whether the drawings to be used in later studies with child subjects are truly ambiguous, or instead call out such a narrow range of potential interpretations as to form something more akin to Rorschach “populars.” Further, do adults recognize that no one could be expected to know what the underlying picture is based on the cryptic set of lines visible through the viewing window? And, finally, do they understand that it is unlikely for two persons asked to comment on the drooodle to offer the same interpretations? With the results of these background efforts are in hand, attention is then turned to the reactions of young school-aged children for the remaining five studies.
Study 2 demonstrates that merely presenting subjects with ambiguous stimuli does not allow a simple discrimination between children who do and do not possess an interpretive capacity. Rather, one must arrange circumstances such that the beliefs a subject ascribes to the target characters are clearly understood to be equivalently mistaken, yet somehow equally legitimate (but still false) beliefs. This is accomplished by comparing the performance of children on two versions of the droodle task: one version in which the child sees only the restricted view and the child is asked about the thoughts of two puppets who have also seen only a restricted view, and another version that gives the child privileged access to the contents of the larger drawing. Having demonstrated that these stimuli and procedures are equal to the task of sorting out those children who do and do not possess the beginnings of an interpretive theory of mind, the remaining four studies are used to first replicate and extend a set of earlier findings (Study 3), then to conduct a direct test of the main hypothesis concerning the relation between false belief understanding and interpretive capabilities (Study 4), and finally to rule out certain alternative accounts of these results (Studies 5 and 6). The logic of this study sequence is described in greater detail in the sections that follow.
STUDY 1

As discussed above, droodles would only be useful for the purpose of assessing children's interpretive capabilities to the extent that they could first be shown to be legitimate objects of interpretation among persons who can be safely assumed to already have an interpretive theory of mind. That is, unless adults whose interpretive capabilities are somehow beyond question, are prompted to provide more than one interpretation of such stimuli, or otherwise give clear evidence of seeing that such drawings are meant to afford multiple interpretations, there would be little real point in presenting droodles to young children. Study 1 is meant to accomplish this background task.

In submitting the droodle stimuli to a panel of adult subjects, several objectives were satisfied. First and foremost among these was to determine if adults do, in fact, respond to the stimuli with multiple interpretations. In other words, do adults see that these stimuli—like clouds or inkblots—can be assigned many different meanings? Secondly, and from a more practical standpoint, it was essential that the stimuli to be employed in subsequent studies with children be rarely interpreted in exactly the same way by different subjects. The danger to be avoided here is the selection droodles that result in "populars" or interpretations that are so common as to offer little real latitude for interpretation. If that were to be the case, we would risk classifying as "interpretive" those non-interpretive children who offered an alternate reading of that stimuli for no other reason than that they were simply reading an unintended message formed by the particular droodle view presented.

There is also the still open question of whether or not adults would immediately realize that multiple interpretations of the stimuli are possible. That is, while this capacity is surely ever-present in principle, its application may be reserved for occasions when more is 'at stake' than the likely contents of a picture. Also, the scoring of children's responses in later studies is meant to be sensitive to "contamination" and "reality" errors in which some feature or the actual identity of the underlying picture is incorporated in the response. In scoring responses to the ship/witch droodle, for example, marine themes are viewed with suspicion. An analysis of the responses of adults who
have not seen the full underlying picture would provide a base rate for these problematic response categories. In Study 1, a sample of adults were presented with the restricted views of the pictures used in the later studies and asked what two hypothetical others—Raggedy Ann and Andy, for the sake of continuity—who had not seen these pictures before, might think of the pictures. Two separate conditions were used. In the first condition, a sample of adults saw both the larger picture and subsequently the droodle, or restricted, view, while in the second, another sample of subjects saw only the restricted view. This second condition is included as a way of determining whether adults need to be prompted into generating more than one interpretation by having access to a "reality" that is denied the puppets.

The purpose of Study 1 was to examine the responses of adult subjects to the droodle stimuli with particular attention to the following data:

1) The frequency with which adults in each condition report that Andy and Ann would give the same responses (i.e., the number of adults who fail to be interpretive in this testing context).

2) The frequency with which adults give responses that would be considered "contaminated" or classified as reality errors using the scoring criteria that will later be applied to children's responses when the adults either do or do not see the full underlying picture (Conditions 1 & 2).

3) The variability of responses elicited by individual stimulus drawings (i.e., the number of distinctly different responses each drawing generates).

4) The effect of varying the amount of information that is available to the subject relative to the puppets (i.e., do adults need privileged knowledge about the "true" contents of the full or unmasked drawing in order to act interpretively).

The time-honoured approach to acquiring data from adult subjects is to test a sample of undergraduate students. In Study 1 an alternative recruitment strategy was used: subjects were recruited and tested "on the Internet." Messages were sent to various Internet news groups that are frequented by persons interested in psychology (for example, sci.psychology.theory) asking that the reader use the World Wide Web to view the stimuli in an electronic questionnaire format. While this approach has the obvious disadvantage of rendering one's subjects somewhat more
anonymous than a traditional in-class testing session, basic demographic data can still be gathered. One clear and compelling practical advantage of this format is that it provides data that is already transcribed and can be easily imported by database software.

Subjects

Subjects were recruited through a posting sent to psychology-related USENET news groups. A copy of the posting appears as Appendix I. Respondents were asked to provide only their age and gender. A total of 205 responses were received. Of these, 114 identified themselves as male, and 96 as female. Age ranged from 21 to 57 years ($M=36.01$, $SD=9.58$) with male respondents being slightly older ($M=39.0$, $SD=10.67$) than females ($M=33.1$, $SD=7.24$).

Materials

The experimental materials consisted of a set of hypertext documents made available through the World Wide Web. Hypertext documents allow Internet users to view documents composed of text and graphics that are stored at remote sites (in this case, on a computer in Vancouver). A print-out of a portion of the questionnaire appears as Appendix II. For the foreseeable future, the full on-line questionnaire (beginning with the Consent Form) can be viewed on the Internet at the following location: <http://vanbc.wimsey.com/~chrisl/Help.html>.

Design & Procedure

The initial USENET posting directed potential subjects to an introductory hypertext document stored in my Internet account that invited them to participate in the study by completing an on-line questionnaire. If they chose to continue, they were required to view a consent form before the questionnaire was displayed.

The on-line questionnaire consisted of a ‘page’ containing a short set of instructions, the stimuli (those labeled Elephant, Cow, Witch, Giraffe, Pooh, Tower, and Shark, in Appendix III),

---

4 Because the observed male-to-female ratio (1.12:1) was in marked contrast to accepted estimates of 1.8:1 among current Internet users, plans to archive responses from “surplus” males were abandoned.
and a set of text boxes in which responses could be typed (see Appendix II for a reproduction of a sample item). As a precaution, subjects were also asked for their first name and e-mail address in the unlikely event that a transmission problem truncated their responses. After completing the questionnaire, the subject could elect to submit their responses to me via e-mail by clicking on a button labeled “Submit Responses.” A computer script ensured that the questionnaire was complete (i.e., all items were filled in) and that the required demographic information had been provided. After submitting their responses, subjects saw a new page that thanked them for their participation and offered them the opportunity to provide contact information if they wished to receive a summary of the results of the study (73 of those who completed the questionnaire asked for a summary, along with 13 who did not submit responses).

Two different versions of the questionnaire were used. Subjects who completed Version A saw only the restricted views of the seven droodles. In Version B, both the restricted view and a small version of the larger picture were displayed. Access to the files that contained these versions was alternated on the web site to ensure that an equal number of subjects completed each version of the questionnaire. The e-mailed responses were imported into a database file for coding and analysis as outlined below.

Results

Classification of Responses

The subjects’ first response to each droodle (i.e., the response they attributed to Ann) was coded for false belief understanding according to the following criteria: if the attribution had no obvious connection to the underlying picture (e.g., if the subject responded that Ann would think the “ship/witch” droodle was “an arrow”), it was coded as False Belief. If the response contained an obvious reference to the underlying picture (e.g., “A ship and a witch’s hat”), it was coded Reality Error. If the response made no explicit reference to the underlying picture, but nonetheless could be seen to have a direct connection to the larger scene (e.g., “A boat about to hit an iceberg”), it was judged a Contamination Error. These criteria were applied to both versions of the
questionnaire—that is, it was possible to commit a *Reality Error* even in Version A where the subject had not seen the underlying picture since the subject could correctly guess the contents or may have previously seen the picture elsewhere. It should be stressed, however, that it is actually a misnomer to use the term “false belief” to describe responses from subjects who never see the full picture.

Pairs of responses—that is, the responses attributed to Ann and Andy for any particular droodle—were classified as *Interpretive* if the responses given for Ann and Andy were scored as *False Belief* according to the criteria listed above, *and* if the responses were different from one another. Thus, to qualify as an *Interpretive* response pair, the subject was required to provide two different false beliefs for a particular droodle, for example, “Ann would think this is an arrow” followed by “Andy would think this is two knife points.” If a response failed to satisfy these criteria, it was classified as *Non-Interpretive–Repetitive* if the responses given for Ann and Andy are identical, or *Non-Interpretive–Mixed* if the pair included non-identical responses where one could be classified as either *Reality Error* or *Contamination* by the criteria listed above.

The number of unique responses provided for each droodle was also calculated as a way of identifying droodles that might prompt only a restricted set of alternative interpretations.

**False Belief Attributions**

Subjects were awarded a single point for each belief attributed to Ann that was coded as *False Belief* according to the criteria described above. Scores could range from 0–7. The mean score for subjects completing Version A of the questionnaire was 6.52 (*SD*=.79), while those who completed Version B had a mean score of 6.24 (*SD*=1.31). An analysis of variance showed that this difference constituted a statistical trend toward lower scores among those who completed Version B (*F*[1,208]=3.635, *p*=.058). These high overall scores (1350 correct of a possible 1470) left little room for differences between the groups in terms of errors. Of the errors that were committed, those from subjects who completed Version A were predominantly *Reality Errors* (50 of 55 errors) while those who completed Version B were more likely to commit *Contamination*
Errors (35 of 65). This pattern seems largely due to “lucky guesses” on the part of those completing Version A, and conservative scoring of responses from Version B (e.g., “a vacuum sucking up a ball” for the Elephant droodle was judged Contamination Error if it arrived from Version B, but False Belief if it arrived from Version A).

A second rater scored the responses from 30 subjects (15 from Version A and 15 from Version B) drawn at random. In addition, the second rater recoded all individual responses that were judged to be errors. Among the 210 (30 subjects each responding to 7 stimuli) responses that were coded by both raters, 5 disagreements were noted. In each case these involved responses originally judged to be Reality Errors that were coded Contamination by the second rater. Of the 120 individual errors that were recoded, the raters agreed on the scoring of all but 18 responses. Again, all disagreements involved responses originally coded as Reality Errors that the second rater scored as Contamination. All coding disagreements were resolved through negotiation.

**Interpretive Attributions**

Subjects were awarded a single point for each response pair judged Interpretive (that is, each time they attributed different false beliefs to Ann and Andy for a single droodle). The mean score for those completing Version A (M=5.14, SD=1.17), was reliably higher than that for Version B (M=4.29, SD=2.13, F[1,208]=13.06, p<.001). Errors were also distributed differently: those completing Version B were not only more apt to make errors, but they were more likely to attribute the same belief to Ann and Andy than were those completing Version A.

The second rater also examined the pairs of responses given for each stimulus within the set of 30 drawn for recoding of false belief described above. Full agreement between the raters was achieved in all but 3 cases in which the second rater judged the items that made up the pair to be too similar to be judged Interpretive.

**Stimulus Suitability**

The number of unique responses for each of the 7 droodles was calculated from the responses attributed to Ann. This offered a maximum of 210 different attributions per droodle. Each droodle
managed to amass at least 25 unique responses, although each had a set of common themes. For the Witch droodle, the most common responses were variations on either teeth of some kind, or shark/dolphin fins, in addition to all manner of “sharp” things. For the Giraffe, bugs, snails, and trees were the most popular attributions, and for the Cow it was globes/planet/earth and ball. The most popular response for the Tower droodle was road or bridge, for Pooh it was mountain and penis. For the Shark, boot and chair, and for the Elephant, it was hockey stick and puck. Each of these responses accounted for more than 10 but less than 25% of the total responses received for each droodle.

Discussion

For the most part, the adult subjects saw the stimuli as an opportunity to showcase their interpretive skills. That is, the lion’s share of individual responses (92%) were classified as “False Belief,” and a smaller, but still sizable proportion of response pairs (70%) were classified as “Interpretive.” When reality errors appeared at all, they were more common in Version A where the subjects never saw the underlying picture, and so, constitute either lucky guesses or prior familiarity with the stimulus picture. This was possible since, in addition to the two droodles published by Price (Witch and Elephant), three others (Shark, Pooh, and Cow) were taken from cartoons or children’s books. In the comments submitted with their responses, several subjects admitted to having seen one or another of these pictures previously. Similarly, while contamination errors were expected to be rare since one assumes these subjects have a firm grasp of false belief, such errors were, as predicted, more common in Version B where subjects had access to the larger picture.

The lower Interpretive scores from respondents to Version B could indicate adults find it more difficult to behave interpretively when they know the contents of the underlying picture than when they do not, and, while one might expect that all of these subjects (and perhaps especially those who viewed Version B) would have perfect False Belief and Interpretive scores, there are several factors that conspire against such ideal data. First, we should not expect adults to be fully interpretive in each and every situation—after all, we sometimes talk to and kick our cars. Second,
and perhaps more importantly, it appears that many of the respondents took it upon themselves to respond as though they were children themselves, or on behalf of the puppets whom they took to be children. Finally, many respondents seemed to assume (and many noted this in their comments) that since the puppets were named “Ann” and “Andy” there must be some undercurrent of interest in sex differences. This led to some rather strange results in their attempts to either maximize or minimize evidence of sex differences in their answers and seems to account for a substantial proportion of cases in which Ann and Andy thought “the same thing.”

Still, the data from this study appear to provide sufficient justification for using the droodle stimuli in later studies with child participants: The adults were largely successful in attributing false beliefs, and different false beliefs about the same stimulus to Ann and Andy, and the individual droodles do not elicit sets of “popular” responses.
**Study 2**

If the aim of this research is to entice children who are capable of doing so into generating multiple interpretations of a single object or to otherwise comment upon differing interpretations of one and the same thing, then, one might wonder, why not simply show such young persons a droodle view and ask what they and another person might think? That would seem to be the simplest case in which one could showcase an interpretive capacity. There are good reasons to suspect, however, that this is not an adequate test of children's interpretive abilities. It is entirely possible (as the findings of Study 1 suggest) that this procedure runs an increased risk of false-negatives. That is, when the larger picture is still an unknown, some children who are fully capable of behaving interpretively, may assume that this is a task of discovering what the picture "really is," and, rather than reporting on a set of possible candidates, would merely announce what they considered to be the "right answer." It is only when they have reason to believe that the real fact of the matter has already been decided—when they have seen the underlying picture, but others have not—that it becomes more compelling to attribute false beliefs or legitimate interpretations to others.

Alternatively, children may approach the task as one of fielding a series of guesses regarding the contents of the window, all the while still believing that it is a task of discovery. In this case, their multiple guesses could be misinterpreted as multiple interpretations. In the procedure followed here, multiple responses can only be confidently counted as interpretive when there is some real fact of the matter at stake and more than one non-factual response is possible. When the true contents of the picture are unknown, these conditions are not met. While simply presenting the droodle views from the outset, then, has the obvious attraction of simplicity, it lacks the potential discriminative power suggested in earlier studies that used these stimuli (Chandler & Helm, 1984; Taylor, 1988). To remedy this shortfall, one could go on, after having secured attributions for the droodle window alone, to show the child the full picture and then repeat the procedure in the more usual manner. This would help settle the question of whether or not subjects who had it within them to behave interpretively, actually require the more convoluted testing procedures (such as
those used in Studies 3-6, or whether simpler procedures will do just as well. In other words, this second study was meant to demonstrate that, when trying to assess interpretive capabilities, there is an important distinction to be drawn between, on the one hand, saying more than one thing about a droodle by simply guessing, and, on the other, having more than one thing to say about how the same stimulus can be differently interpreted.

Subjects
A total of 68 children (39 boys, and 29 girls) were recruited for participation from after-school care facilities in the Vancouver area. Two subjects failed to complete testing: one withdrew her consent, and testing of the other was disrupted by other children. Data reported here are for the remaining 66 subjects. All subjects (39 boys, 27 girls) were tested between their 5th and 7th birthdays. Ages ranged from 60–84 months \((M = 71.8 \text{ mos.}, SD = 6.98)\). Subjects were divided for reporting purposes into groups of 5- and 6-year-olds. The group of 5-year-olds (mean age = 65.0 mos., \(SD = 2.68\)) included 18 boys and 12 girls. The group of 6-year-olds (mean age = 77.1 mos., \(SD = 3.85\)) included 21 boys and 17 girls. Although not all subjects were native speakers of English (18 of 68 were not), all of these were judged by the staff of the facility to be competent in their comprehension and production of the English language.

Materials
Four of the seven droodles employed in Study 1 were used in Study 2 (stimuli labeled Giraffe, Snake, Shark, and Tower in Appendix III). Because at least some of the adult subjects in Study 1 had previously seen three of the droodles (two drawn from Price’s original set, and one that was constructed specifically for this procedure), those stimuli were not used in this procedure where the subjects were presented with only the restricted view. Single window versions of the droodles were used.

Design & Procedure
A total of four trials were conducted: two Restricted View trials, followed by two Unrestricted View trials. In Restricted View trials, the subject saw the restricted, or droodle view and never had
access to the full picture. In Unrestricted View trials, the subject first saw the full underlying picture followed by the droodle view. The order of these conditions or trial types was not counterbalanced since presenting an Unrestricted View trial would likely have led subjects to expect the presence of a larger underlying picture in the Restricted View trials that might follow. To reduce this possibility even further, the stimuli used for the Restricted View trials were constructed without covers—that is, they were droodles in the manner produced by Price and consisted of a single sheet of paper bearing what, in other presentations, would be a restricted view of a larger picture.

Subjects were tested individually in a quiet room within their after school care facility (see Figure 6, below). Subjects were seated at a small table and introduced to two 30 cm dolls ("Raggedy Ann" and "Raggedy Andy") then asked to "pretend that these dolls are real people like you and me." Two cardboard boxes served as the dolls’ "houses" and subjects were told that "when Ann and Andy are inside their houses, they can’t hear what we’re saying, and they can’t see what we’re doing." The dolls were then placed inside their respective houses, and the subject was asked “Can Raggedy Ann hear us talking right now?”, and “Can Raggedy Ann see us right now?”

Figure 6: Experimental set up for Study 2.

Subjects were shown the first of the Restricted View droodles and told: “Raggedy Ann and Raggedy Andy have never seen this picture before. Let’s get Ann out of her house and show her

---

5 A total of 12 subjects (all boys) were recruited from a baseball team. These subjects were tested on the sidelines during a team practice. An additional 5 subjects were tested in their homes.
this picture.” The child then retrieved Ann from her house. “Now, Ann has never seen this picture before. What will Ann say this is?” The child’s response was noted and repeated aloud by the experimenter. This same procedure was then repeated for Andy: “Let’s get Andy out of his house and show him this picture. Now, Andy has never seen this picture before. What will Andy say this is?” The child’s response was again be noted and repeated aloud. A second Restricted View trial immediately followed.

Unrestricted View trials began by showing the child the full picture and asking him or her to describe the picture. If the child’s response failed to take into consideration those particular sections of the picture that would later be visible through the window in the cover (for example, the bow of the ship or the witch’s hat), they were asked to specifically comment on those features of the picture. The child was then told: “Raggedy Ann and Raggedy Andy have never seen this picture before. Let’s get Ann out of her house and show her this picture. But we’re going to show her the picture like this [cover is closed to reveal only the restricted view]” The child then retrieved Ann from her house. “Now, Ann has never seen this picture before. What will Ann say this is?” The child’s response was noted and then the procedure was repeated for Andy: “Let’s get Andy out of his house and show him this picture. Now, Andy has never seen this picture before. What will Andy say this is?” This same procedure was repeated for the second Unrestricted View droolde.

Results

SCORING OF INDIVIDUAL RESPONSES

Separate analyses were conducted for the Restricted View and Unrestricted View trials. In both cases, scores for False Belief understanding were derived from responses given for the first puppet who viewed each stimulus picture. Scores for Interpretive understanding were derived from the response pairs given for the two puppets to each stimulus picture.

6 The order of the puppet’s appearance (that is, who saw the picture first) was counterbalanced across subjects rather than trials to avoid unnecessarily confusing either the subject or the experimenter.
In Restricted View trials all responses should qualify as False Belief according to the criteria used in Study 1, since there is no "reality" that might otherwise intrude into the responses offered on behalf of the two puppets. Similarly, Reality and Contamination errors cannot be made unless the subject has some prior familiarity with the drawing. Since the three droodles that gave some slight evidence of having been familiar to some adult subjects in Study 1 were not used in this study, these three scoring categories may seem irrelevant to the Restricted View trials. Still, these categories were included for two reasons. First, because inter-rater reliability was determined by asking a second rater to examine data from both a sample of subjects as well as a sample of responses, it was important to keep that rater blind to the condition in which the responses were produced. Second, it seemed prudent to guard against subjects' prior exposure to the stimuli in this study as well.

Given that False Belief scores on Restricted View trials were effectively guaranteed to be at or near ceiling, the question of real interest with respect to these trials was whether or not the subjects attributed different beliefs to Ann and Andy. As noted above, however, attributing different responses to the puppets on these trials may not indicate an appreciation of the interpretative character of knowing, since the subject may merely be placing their own "guesses" into the mouths of Ann and Andy, or assigning one puppet a "right" answer and the other a "wrong" answer.

On Unrestricted View trials, when the subject knows the contents of the underlying picture, responses can be scored as described in Study 1 (i.e., False Belief, Reality Error, Contamination Error). As above, the child's attribution to a puppet was scored False Belief if their attribution made no reference to the underlying picture. If the attribution was simply a calling out of what the underlying picture actually is, it was scored Reality Error. Less obvious errors, those that somehow contain features of the underlying picture were labeled Contamination Error. Subjects' understanding of false belief was estimated by giving them a single point for each False Belief response on the two Unrestricted View trials, thus scores varied from 0 to 2. Subjects with summary scores of 0 were classified as False Belief Understanding Absent, those with scores of 1 as False Belief Understanding Transitional, and scores of 2 as False Belief Understanding Present.
A second rater coded the responses from 20 subjects chosen at random. The second rater also coded all individual responses that were originally coded as Contamination Errors. Agreement between raters was observed on 73 of the 80 responses coded. The 7 disagreements all concerned responses that were originally coded Reality Error but which the second rater scored as Contamination. Of the remaining 11 Contamination errors, the raters agreed on all but 2 one of which was coded as a Reality Error and the other as False Belief. These disagreements were resolved through discussion.

SCORING OF RESPONSE PAIRS

Subjects' interpretive understanding was estimated by examining their attributions to Ann and Andy on any single trial. A response pair was considered Interpretive if the responses attributed to Ann and Andy for the same picture would each be scored False Belief by the criteria listed above, and the responses were different from one another. All other pairs were classified as Non-Interpretive. Pairs that included a False Belief and Reality Error or Contamination were referred to as Non-Interpretive-Mixed, while those that contained the same belief attributed to both puppets (whether the same false belief or the same error) were referred to as Non-Interpretive-Repetitive. Subjects were awarded a single point for each occasion on which they attributed different false beliefs to the puppets on a single trial. That is, scores could vary from 0–2. Individual subjects Interpretive Understanding was classified as follows: scores of 0 = Non-Interpretive; scores of 1 = Transitional; scores of 2 = Present.

The second rater agreed with all but two of the original codings given to the pairs of responses from the sample of 80 pairs provided. In both cases, the second rater felt that the attributions given for the puppets were too similar to one another.

FALSE BELIEF UNDERSTANDING

Restricted View Trials. Applying the response scoring criteria described above yielded the following results for the Restricted View trials:
Table 3: Classification of Responses for False Belief Understanding (Restricted View Trials by Age Group, Study 2).

<table>
<thead>
<tr>
<th>Response Classification</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Belief</td>
<td>50</td>
<td>69</td>
<td>119</td>
</tr>
<tr>
<td>Contamination</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Reality Error</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Don’t know/No Response</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>58</td>
<td>74</td>
<td>132</td>
</tr>
</tbody>
</table>

As can be seen from this table, 119 of the 132 responses (90.1%) were classified as instances of false belief attribution. Combining the Contamination and Reality Error categories reveals a total of 11 responses (8.3%) that bear some mark of privileged knowledge. That is, the subjects seemed to know the true contents of the larger window, in a way that they should not, given only their restricted view of that underlying picture. In fact, on these trials there was no underlying picture. In each of these cases, however, the responses could be written off as “lucky guesses”—for example, “a shirt” or “fish scales” for the Shark, “a hat” for the Snake, “a tower thing” for the Tower. Judging these to be either Reality Errors or Contaminations, turned out to be a matter of deciding whether the response referred to a large enough part of the underlying picture. In no instance did the subjects give any indication of having seen the picture prior to testing.

Subjects were then assigned a single point for each response classified as False Belief. For the total sample, the mean score (of a possible 2.0) was 1.81 (SD=.40). For 5-year-old subjects, the mean score was 1.72 (SD=.35), and for 6-year-olds 1.86 (SD=.06). The frequency of False Belief scores for each age group appears in Table 4 below.

Table 4: False Belief Understanding Scores (Restricted View Trials by Age Group, Study 2).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>False Belief Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>0</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
</tr>
</tbody>
</table>
Based on these scores, individual subjects were then classified (using the criteria described above) as having an understanding of False Belief that was Absent, Transitional, or Present. By these standards, 53 of the 66 subjects (80.3%) appeared to understand false belief (i.e., are classified as False Belief Present) when both they and the puppets viewed only a restricted portion of the picture. It must be remembered, of course, that the beliefs attributed to the puppets are only “false” from the perspective of persons who know the true contents of the underlying picture.

Unrestricted View Trials. A more legitimate measure of subject’s understanding of false belief comes from an examination of their responses on the Unrestricted View trials in which the subjects are made privy to the contents of the full picture before the cover is closed and the puppets are shown only the droodle view. Here at least the puppets can be legitimately said to have false beliefs. Classifying individual responses using the criteria employed above, yields the results shown in Table 5 below:

| Table 5: Classification of Responses for False Belief Understanding (Unrestricted View Trials by Age Group, Study 2). |
|--------------------------------------|------------------|------------------|------------------|
| Response Classification               | 5-year-olds      | 6-year-olds      | Total            |
| False Belief                         | 40               | 62               | 102              |
| Contamination                        | 6                | 3                | 9                |
| Reality Error                        | 12               | 9                | 21               |
| Don’t know/No Response               | 0                | 0                | 0                |
| Total                                | 58               | 74               | 132              |

Under these more canonical circumstances, when the child subject knows something that the puppets do not, a total of 102 of the 132 responses (77.3%) of all responses qualify as false beliefs. Errors account for all of the remaining responses (22.7%), with the majority of these being Reality Errors in which knowledge of the true content of the picture is ascribed to the puppet. Classifying individual subjects according to the number of False Belief responses produced, results in the data shown in Table 6, which shows that 38 of the 66 subjects (57.6%, 41.4% of the
5-year-olds, and 70.3% of the 6-year-olds) could be classified as *False Belief Understanding Present* (i.e., those with scores of 2).

**Table 6: False Belief Understanding Scores (Unrestricted View Trials by Age Group, Study 2).**

<table>
<thead>
<tr>
<th>Age Group</th>
<th>False Belief Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>1</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>1</td>
</tr>
</tbody>
</table>

Comparing False Belief Scores across Trial Type (Restricted vs. Unrestricted View) and Age Group (5- vs. 6-Year-olds) in a repeated measures ANOVA, revealed reliably lower False Belief scores on Unrestricted View trials ($F[1,64]=9.93, p<.01$, mean scores = 1.54 and 1.80) and reliably lower scores for the group of 5-year-olds ($F[1,64]=6.65, p<.05$, mean scores = 1.55 and 1.77). The Age Group by Trial Type interaction, however, was not statistically significant ($F[1,64]=.893, p=.348$). These results suggest that subjects in general, and perhaps 5-year-olds in particular, find it harder to attribute a 'false belief' when the true contents of the picture are known to them. Since the beliefs attributed in the Restricted View trials cannot be considered legitimate false beliefs, the conclusion to be drawn from these data seems to be that, when there is no truth of the matter that is selectively revealed to the subject but not the protagonist, false belief competence will be overestimated (see Figure 7).

**Figure 7: Mean False Belief scores by Trial Type and Age Group (Study 2).**
**INTERPRETIVE UNDERSTANDING**

*Restricted View Trials.* Applying the scoring criteria for response pairs described above yielded the following results for the *Restricted View* trials:

Table 7: *Classification of Responses for Interpretive Understanding (Restricted View Trials by Age Group, Study 2).*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Interpretive</th>
<th>Non-Interpretive–Mixed</th>
<th>Non-Interpretive–Repetition</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>5</td>
<td>12</td>
<td>41</td>
<td>58</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>18</td>
<td>7</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>19</td>
<td>90</td>
<td>132</td>
</tr>
</tbody>
</table>

By these criteria, just 8.6% (5/58) of the 5-year-olds and 24.3% (18/74) of the 6-year-olds response pairs were classified as Interpretive on *Restricted View* trials. The percentage of response types for each age group are presented graphically in Figure 8 below:

Figure 8: *Percentage of Interpretive Response Types (Restricted View Trials by Age Group, Study 2).*

Assigning subjects a single point for each Interpretive response pair, results in the data shown in Table 8.
Table 8: Classification of Individual Subjects’ Understanding of Interpretation (Restricted View Trials by Age Group, Study 2).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Interpretive Understanding</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td></td>
<td>25</td>
<td>24</td>
<td>49</td>
</tr>
<tr>
<td>Transitional</td>
<td></td>
<td>3</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>29</td>
<td>37</td>
<td>66</td>
</tr>
</tbody>
</table>

Unrestricted View Trials. Scoring the 132 response pairs from trials on which the subjects saw the full picture, but the puppets did not, yields the data shown in Table 9. On these trials, just 10.4% of the response pairs provided by 5-year-old subjects, and 39.2% of 6-year-old’s were classified as Interpretive (see Figure 9).

Table 9: Classification of Responses for Interpretive Understanding (Unrestricted View Trials by Age Group, Study 2).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Response Pair Classification</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretive</td>
<td></td>
<td>6</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>Non-Interpretive–Mixed</td>
<td></td>
<td>19</td>
<td>15</td>
<td>63</td>
</tr>
<tr>
<td>Non-Interpretive–Repetition</td>
<td></td>
<td>33</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>58</td>
<td>74</td>
<td>132</td>
</tr>
</tbody>
</table>

Figure 9: Percentage of Interpretive Response Types (Unrestricted View Trials by Age Group, Study 2).

Table 10 shows the number of subjects in each category when individual subjects are classified by their Interpretive Understanding scores.
Table 10: Classification of Individual Subjects' Understanding of Interpretation (Unrestricted View Trials by Age Group, Study 2).

<table>
<thead>
<tr>
<th>Interpretive Understanding</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>25</td>
<td>21</td>
<td>46</td>
</tr>
<tr>
<td>Transitional</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Present</td>
<td>2</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>37</strong></td>
<td><strong>66</strong></td>
</tr>
</tbody>
</table>

Comparing Interpretive Scores across Trial Type (Restricted vs. Unrestricted View) and Age Group (5- vs. 6-year-olds) in a repeated measures ANOVA, revealed reliably higher Interpretive scores on Unrestricted View trials ($F_{[1,64]}=10.99$, $p<.01$, mean scores = .530 and .348) and reliably higher scores for the group of 6-year-olds ($F_{[1,64]}=6.88$, $p<.02$, mean scores = .635 and .190). A significant interaction between Age and Trial Type ($F_{[1,64]}=5.66$, $p<.03$) indicates that these results are entirely due to the performance of the 6-year-old subjects whose scores were reliably higher on the Unrestricted View trials (.784 vs. .486), while the scores of the 5-year-olds were not affected by this manipulation (mean scores = .207 vs. .172). In contrast to the findings reported above for false belief understanding, the take-home message from these analyses of subjects' interpretive understanding seems to be that the Restricted View trials underestimate subjects' understanding of interpretation (see Figure 10).

Figure 10: Mean Interpretive Scores by Trial Type and Age Group (Study 2).
COMPARING FALSE BELIEF AND INTERPRETIVE UNDERSTANDING

As noted above, it is apparent that Restricted View trials provide a non-interpretable and seriously inflated estimate of the False Belief competence of the subjects, while largely underestimating their Interpretive competence (see Figure 11, below). For these reasons, in the remaining analyses, only scores from the Unrestricted View trials were used.

*Figure 11: Mean False Belief and Interpretive Scores by Trial Type (Study 2).*

As a way of comparing individual subjects' performance on measures of false belief and interpretive understanding, a contingency table was constructed that assigned each subject to a cell based on their False Belief and Interpretive classifications on the Unrestricted View trials (see Table 11). Several of the cells in this table are shaded to indicate that they will not be used in subsequent analyses. These cells are excluded for one of two reasons. Three cells are empty by definition according to the scoring procedures: False Belief Absent, Interpretation Transitional; FB Absent, Int. Present; and FB Transitional, Int. Present (i.e., row 1 column 2; row 1, column 3; row 2, column 3). These 3 cells could not contain observations because, to be classified as an Interpretive response pair, each pair must begin with a response that has already been classified as False Belief. Thus, a child who gave no responses that were classified as False Belief could not attain anything other than an Interpretation Absent classification. Similarly, since all of one's responses would need to be classified as False Belief in order to make it into the running for Interpretive Present, no False Belief Transitional subject could ever appear in the Interpretive
Present column. The cells labeled Absent/Absent and Present/Present (i.e., row 1, column 1; row 3, column 3), will be excluded for the simple reason that these cells contain those developmentally uninteresting subjects who have either all or none of the two capabilities under consideration. In other words, since I am interested in determining the degree of overlap between measures of false belief and interpretive understanding, subjects who are either at ceiling or floor on both measures offer us little in the way of real guidance. While this eliminates one quarter of the current sample, the remaining 49 subjects do allow a direct test of the hypothesis that false belief understanding is not sufficient for an appreciation of interpretation.

Table 11: Contingency Table of Subjects’ False Belief and Interpreive Classifications (Study 2).

<table>
<thead>
<tr>
<th>Understanding of False Belief</th>
<th>Understanding of Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absent</td>
</tr>
<tr>
<td>Absent</td>
<td>2</td>
</tr>
<tr>
<td>Transitional</td>
<td>25</td>
</tr>
<tr>
<td>Present</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
</tr>
</tbody>
</table>

A Chi-Square analysis of the resulting 2x2 matrix did not support the notion that these are redundant measures, however ($\chi^2[1, N=49] = 2.44, p=.118$), nor did a subsequent analysis that restricted attention to the 6-year-old subjects alone, although this smaller sample came close to indicating a statistical trend ($\chi^2[1, N=23] = 2.65, p=.103$).

Chi Square analyses were conducted in addition to the Analyses of Variance, in order to focus attention on those subjects who held some potential for revealing a distinction between the measures of false belief and interpretive understanding. In other words, by weighting the analysis such that it was particularly sensitive to the subset of subjects who might have been classified as Transitional on both measures, it was possible to examine this relation with greater precision. While these analyses may be limited in terms of their individual statistical power, the effects will be shown to be repeatable across studies and an overall analysis covering several studies and a much larger number of cases will be reported toward the end of the document. Where appropriate, all subsequent Chi Square analyses with 1 degree of freedom will be reported with Fisher’s Exact corrections where expected cell values fall between 1 and 5. Yates corrections were not used following Grizzle’s (1967) finding that it constitutes a “dangerously conservative test” and expected values were allowed to fall to 1 on the advice of Remington and Schork (1970) and Cochran (1954). Finally, with reference this particular analysis, I know that I shouldn’t be reporting a probability value above .10, but who could resist $p=.103$?
Discussion

The purpose of Study 2 was to show that in order to estimate a child’s grasp of the interpretive possibilities that are inherent in the droodles stimuli, one must first reveal the larger picture of which the droodle is a part. This demonstration was intended as a way of empirically underlining the fact that careful attention needs to be paid to the truth values that subjects’ assign to their attributions to the puppets. For children who are already capable of ascribing to different persons different interpretations of one and the same stimulus, it simply does not matter whether either or both of these interpretations are judged to be right or wrong. In contrast, for children who lack the capacity to view the knowing process interpretively, it is critical to know when they are assigning a true belief to one protagonist and a false belief to another. This would reveal only an understanding of false belief—not interpretation—since simply saying two different things when asked what first one, and then another person thinks, is not sufficient to demonstrate an understanding of interpretation. It tells us something about the subjects’ grasp of false belief, but nothing about that individual’s competence to consider the prospect that two different persons may differently and legitimately interpret (or misinterpret) one and the same thing.

The procedural route chosen for this demonstration was to contrast the same subjects’ attributions to the puppets under two different testing conditions. In the first, the subject and the puppets remained equally ignorant of the true state of affairs. In the second, the subject was given privileged access to the real truth of the matter while the puppets were kept strategically ignorant. In the first condition, all attributions made to the puppet should technically qualify as “false beliefs” since the subject is unaware of the larger picture. Over 90% of responses given under these circumstances were scored as False Belief (the remainder being ‘lucky guesses’ conservatively scored, and ‘don’t know’ responses). When asked what the second puppet might think of this same picture, however, two thirds of the responses (68.1%) were a simple repetition of whatever had been attributed to the first puppet and just 8.6% constituted a new and different belief. In the second condition, when the subject knew the contents of the larger picture, but the puppets did not, the percentage of responses scored False Belief fell from 90 to 77.3% and when asked what the
second puppet would think, repetitions fell from 68 to 25.7% while the frequency of new and different beliefs rose from 8.6 to 26.5%.

This pattern of findings suggests that simply asking children to attribute beliefs to others about some ambiguous stimulus will result in an overestimation of their understanding of false belief and an underestimation of their understanding of interpretation. It seems likely that on the first set of trials, the subjects viewed the task as one of discovering what the picture “really is” and, having satisfied themselves that they had done so, merely repeated the same belief for both puppets. In later trials when the truth had been revealed to them, subjects struggled to deny their own privileged knowledge to the puppets and (for those able to do so) worked to come up with different mistaken readings of the same stimulus for each puppet.

A secondary aim of Study 2 was to show that subjects capable of expressing an understanding of false belief in this testing context should not be automatically assumed to also appreciate the interpretive possibilities inherent in the stimuli. Although a statistically reliable difference was not found, the data tend in the predicted direction, with the vast majority of children giving evidence of understanding false belief (64 of 66) but substantially fewer (20 of 66) showing some appreciation of interpretation. Of the 38 who were at ceiling on the false belief measure, half had scores of zero on the interpretive measure and just 15 evidenced a solid grasp of interpretation.

Having established both the suitability of the stimuli (Study 1) and the sensitivity of the procedure (Study 2), attention is turned in Study 3 to certain earlier attempts to chart the course of children’s perspective-taking skills that employed somewhat similar materials and techniques to the procedures used in the present studies.
STUDY 3

The data from Taylor (1988) suggest, but do not explicitly demonstrate, that young children can appreciate that a puppet who sees only a restricted view will hold a false belief about the contents of the picture. In discussing the results of her second experiment, Taylor notes that: “Over half of the 6-year-old children spontaneously speculated on alternative hypotheses for the identities of the parts in the restricted views. For example, children observed that someone might think the rabbit was a boomerang, submarine, airplane, or planet” (p.717). In Study 3 this claim is put to explicit test by repeating Taylor’s procedure and asking the subjects, not a true/false test question (“Will [puppet’s name] know this is a ____?”) as Taylor did, but an open-ended question: “What will [puppet’s name] think this is a picture of?” While the Taylor procedure does not allow us to measure subjects’ understanding of interpretation, this can be easily remedied by asking what a second puppet would think of the same restricted view. This much could work to shore up the findings of Study 2 and help sort out the relation between interpretive understanding and conceptual perspective taking. In addition, Taylor found that training trials with restricted views had a positive effect on 4-year-old subjects’ ability to predict that a puppet would not know the contents of a picture based on seeing only the restricted view. In Study 3 this manipulation was replicated with the addition of a second puppet to determine whether training improves only subjects’ grasp of the possibilities for false belief inherent in the droodles stimuli, or also aids them in their efforts to report upon multiple interpretations.

An examination of Taylor’s data indicates that 5- and 6-year-olds have the most room to maneuver. That is, while 4-year-olds needed to be coaxed into viewing her version of the droodles task as one of assigning false beliefs to the puppet, 8-year-olds perform at ceiling (Taylor gathered no data from 7-year-olds). This, together with the findings from the completed studies in the present sequence which show that 7-year-olds are approaching ceiling on interpretive tasks, lend support for the assumption that 5- and 6-year-olds stand to benefit most from a training intervention.
The purpose of this study is to compare the performance of groups of subjects under three training conditions. The first acts as a control condition for the other two by offering children no special training or instruction. In the second (False Belief Training), subjects are meant to benefit from a series of trials in which the capacity of droodles to promote false beliefs is illustrated. In the third (Interpretive Training), a series of trials are conducted that showcase the fact that droodles can be interpreted in more than one way—that is, that more than one false belief can be had about any one droodle. As in Study 2, subjects' understanding of false belief and interpretation is also assessed as a test of the main hypothesis that the former is necessary but not sufficient for the latter.

Subjects
The subject pool for this study consisted of 49 children recruited from the ranks of those who completed Study 2. Subjects were selected (based on their False Belief and Interpretive scores) in order to assemble a sample of children who could be expected to benefit from the training component of Study 3. Subjects classified in Study 2 as False Belief Transitional or Present, and as Interpretive Absent or Transitional were asked to participate in Study 3. This eliminated children with no evident grasp of false belief as well as those whose understanding of interpretation was at ceiling. One child failed to complete testing in Study 3. Data are reported on the remaining 48 subjects. As reported in Study 2, although not all of these subjects were native speakers of English (9 of 48 were not), all were judged to be competent in their comprehension and production of the English language by the staff of the facility who were familiar with the children.

Subjects ranged in age from 60 to 82 months ($M=70.1$ mos., $SD= 6.48$). Nineteen girls ($M=71.6$, $SD=6.56$), and 29 boys ($M=69.2$, $SD=6.35$) were divided into groups of 5-year-olds ($N=26$, $M=65.0$ mos., $SD=2.67$) and 6-year-olds ($N=22$, $M=76.23$ mos., $SD= 3.74$) and then assigned to one of the three conditions in a way that produced 3 groups of 16 children and a rough equality of age and sex across conditions (see Table 12). Attempts to attain fully equal groups of boys and girls were hampered by an inability to access a larger pool of subjects for Study 2.
Table 12: Distribution of subjects across Conditions by Age and Sex (Study 3).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Control</th>
<th>False Belief</th>
<th>Interpretive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>5 yrs</td>
<td>6 yrs</td>
<td>5 yrs</td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Materials

A set of 8 line drawings were used (those in Appendix III labeled, Elephant, Cow, Pig, Boot, House, Witch, Pooh, and Ship). Three of these (Pig, Boot and Ship) were identical to the stimulus drawing taken from Taylor’s (1988) report (see Figure 3, pg. 30). One picture (House) was constructed from a description given by Taylor (1988) and one picture (Cow) was used twice although different viewing windows were used. Single window restricted views were used in all cases. A small cassette recorder was used to record subjects’ responses for transcription.

Design & Procedure

Children were tested individually in a quiet room. All subjects were tested within 7 days of having completed Study 2, with the majority of subjects tested on the following day.

The procedure for the Control Condition was taken directly from Taylor (1988). The False Belief Training Condition procedure replicated Taylor’s Training Condition. In the Interpretive Training Condition, which otherwise followed the same script as Taylor’s Training Condition, a second puppet was added while the post-training questions regarding ambiguity were deleted (see below).

In all conditions, subjects were introduced to the two puppets as in Study 2, and told: “We’re going to play a game with these two puppets. We’re going to pretend that they are real people just like you and me. This is Raggedy Ann and this is Raggedy Andy. These are their houses. When they are inside their houses, they can’t see what we’re doing and they can’t hear what we’re saying. Let’s put them in their houses now.” As in Study 2, the puppets were then placed in their respective ‘houses’ and the subjects were asked if Ann could “hear what we’re saying right now”
and if Andy could “see what we’re doing.” In addition, the subjects were told, following Taylor’s procedure, that “We’re going to play a game with Ann and Andy. I’m going to show you a picture, then cover it up so that only a small part is showing, and then have Ann [and/or] Andy try to guess what it is.”

In the False Belief Training Condition, the Boot, Ship, and Pig drawings from Taylor (1988) were used (see Figure 4, pg. 33). The restricted views of these pictures were identical to one another. Subjects were shown one of the stimuli and asked to identify the object. The cover was then fitted over the picture and Raggedy Ann was retrieved from her house and shown the restricted view. Following Taylor (1988), the experimenter spoke for the puppet “using a high-pitched voice” (p. 714). Raggedy Ann then guessed incorrectly, saying for example that the boot is a box, the ship is a boot, and the pig is a boot. When the puppet erred, the experimenter said “No Ann, look” and the cover was removed to reveal the full picture. The cover was then closed and the picture left in full view of the child. With the second and third mistakes, the experimenter pointed out that the restricted views were identical. Following the third picture, the puppet was placed back in her ‘house’ and the experimenter said: “Ann made a lot of mistakes, didn’t she? She thought this [pointing to boot] was a box, but it’s a boot [removes cover]; she thought this was a boot but it’s a ship [removes cover]; she thought this was a ship, but it’s a pig [removes cover].” For the House and Cow pictures, the windows were situated to reveal parts of the underlying picture that made it easy for the puppet to identify the object—the area around the door of the house, and most of the cow’s head. In these cases Ann guessed correctly.

The puppet was then held up to ‘see’ one of the restricted views (Pig, Boot, or Boat only) and, again, the child was asked “When Ann sees just this, does she know what this is a picture of?” After the child answered, the puppet guessed incorrectly for each picture. Following Taylor, the purpose of this manipulation was to “make the child realize that the three restricted views are

---

8 As in Study 2, the order of appearance of the puppets was counterbalanced across subjects rather than trials.
9 It might have been best, at least for my own purposes, if these incorrect attributions were to form a circle—that is, if the guess were always one of the 3 ‘truths.’ However, since the word “box” appears more than once in
identical" (p.714). Nine of these trials were conducted unless the child answered correctly on three trials in a row. This much replicated Taylor’s Training Condition.

The Control Condition precisely replicated Taylor’s Control Condition. In this condition the same three Pig, Boot, and Boat stimulus pictures were used but the restricted views had different appearances (see Figure 4, p.26). As above, subjects were shown one of the stimuli and asked to identify the object. The cover was then fitted over the picture and Raggedy Ann was retrieved from her house and shown the restricted view. The experimenter again spoke for the puppet using a high-pitched voice. Raggedy Ann guessed incorrectly, for the Pig, Boot, and Boat pictures, saying that the boot was a ball, the ship was a shark, and the pig was a hand and correctly for the House and Cow pictures. Following this, the puppet was placed back in her ‘house’ and the experimenter presented the Pig, Boot and Boat pictures with covers closed and said: “This is how the pictures looked to Ann. She could only see a little bit of them. She thought this one was a ball, but it’s a boot [remove cover]; she thought this one was a shark, but it’s a boat [remove cover]; she thought this one was a hand, but it’s a pig [remove cover].

The Interpretive Training Condition largely replicated Taylor’s Control Condition with the exception that both Ann and Andy guessed at the contents of the picture. In this condition the same stimulus pictures were used but, as above, the restricted views of the Pig, Boot, and Boat pictures had different appearances (see Figure 4). As above, subjects were shown one of the stimuli and asked to identify the object. The cover was then placed over the picture and Raggedy Ann was retrieved from her house and shown the restricted view. For the Boot picture, Raggedy Ann guessed incorrectly, saying it was a ball. Raggedy Andy was retrieved and guessed that the boot was a leaf. The puppets differed in their guesses for the Pig and Boat pictures was well: Ann said the boat was a shark, Andy said it was a paint brush; Ann said the pig was a hand, Andy said it was a french fry. Both puppets guessed correctly for the House and Cow pictures. Following this, the puppets were placed back in their ‘houses’ and the experimenter presented the Pig, Boot, and

Taylor’s report, and was not, therefore, a typographical error, “box” was used in the trials in Study 3.
Boat pictures with cover closed and said: “Ann and Andy made a lot of mistakes, didn’t they? Ann thought this [pointing to boot] was a ball, and Andy thought it was a leaf, but it’s a boot [removes cover]; Ann thought this was a shark and Andy thought it was a paint brush, but it’s a ship [removes cover]; Ann thought this was a hand, and Andy thought it was a French fry, but it’s a pig [removes cover].” Because the point of these trials was to expose subjects to different interpretations of the same view, and because non-identical restricted views were used in this condition, the subjects were not asked (as they were in the False Belief Training Condition) whether Ann would know the what the picture contained given access to just the restricted view.

Following the training trials, subjects in all three conditions were asked (following Taylor, 1988): “Ann [“Ann and Andy” in the Interpretive Training Condition] made a lot of mistakes, didn’t she [they]? Why do you think she [they] made so many mistakes—because she isn’t [they aren’t] very good at guessing or because the game is too hard?”

Subjects in all conditions then received a total of 4 test trials using the Cow, Elephant, Witch and Pooh stimulus pictures (see Appendix III). On each trial the subject saw the full picture and was asked to identify the object(s) it contained. The child was then told: “Raggedy Ann and Raggedy Andy have never seen this picture before. Let’s get Ann out of her house and show her this picture. But we’re going to show her the picture like this [cover is closed]” The child then retrieved Ann from her house. “Now, Ann has never seen this picture before. What will Ann say this is?” The child’s response was noted (and repeated aloud) and then the child was told: “Let’s get Andy out of his house and show him this picture. Now, Andy has never seen this picture before. What will Andy say this is?”

Two notable departures were made from Taylor’s (1988) original procedure. First, subjects were not given any “Pretraining Trials.” According to Taylor, these trials were designed “to help children appreciate that the views of the pictures shown through the windows were restricted” (p. 706). Since all of the subjects in the present study had previously participated in Study 2 and because all had given good evidence (by way of their False Belief scores) that they understood this concept, such pre-training seemed superfluous. Secondly, no post-test “Recognition Task” was
conducted in which subjects were asked to select the restricted views that were used in the testing phase from an array that included views that revealed either more or less of the underlying picture than did the restricted views used in testing. In Taylor’s original study most of the 6-year-olds were at ceiling on this measure while 4-year-olds tended to pick the views that revealed more of the picture (no 5-year-olds were tested). Since there were no reliable differences between Taylor’s Training and Control groups, there seemed little point in trying to detect the differences between 5- and 6-year-olds using this measure.

Results

In the False Belief Training Condition, all but 2 of the 16 subjects correctly answered 3 times running that Ann would not be able to identify which of the Pig, Boot, or Boat pictures lay beneath the restricted view. Two especially observant subjects detected a slight difference in the restricted view of the Pig picture which they felt Ann could use to identify that picture. This manipulation was meant to assess the extent to which children in this condition realized that since the 3 restricted views were identical, a naive observer seeing only the restricted view could never know for certain which of the 3 larger pictures lay beneath the cover. To that extent, it was largely successful since even the two children who detected a difference in one of the pictures agreed that Ann would not know whether the other two restricted views belonged to the Boot or Boat picture. This suggests that, like the subjects in Taylor’s Study 2, children in the present study were able to “learn that restricted views in general can be ambiguous ...[and]... that a naive observer might not guess the identity of a partially covered object” (p.716).

When subjects were asked to assign blame for the puppets’ failures following the training phase of the study, 12 of 48 subjects (25%) claimed the puppet(s) wasn’t very good at guessing, while the remainder claimed the task was too hard. This pattern is substantially the same as that found by Taylor (1988) for 6-year-olds, with 60% of her subjects blaming failure on task difficulty. The only departure from this general trend in Taylor’s findings was that some two-thirds of the 4-year-old subjects in her Control Condition (11 of 16) blamed the puppet—a finding that was not replicated here with the 5-year-olds, who pointed to the difficulty of the task with even
greater frequency (21/26) than did the 6-year-olds (15/22). No differences were apparent between the training conditions on this measure ($\chi^2[2, N=48] = .667, p = .715$), or between the age groups ($\chi^2[1, N=48] = 1.007, p = .315$, Fisher's Exact $p = .340$).

**FALSE BELIEF UNDERSTANDING**

The data from the testing trials were coded for understanding of false belief using the scoring criteria employed in Studies 1 and 2. Table 13 displays the results of categorizing the responses given for the first puppet to see each stimulus picture as False Belief, Contamination, or Reality Error.

**Table 13: Classification of Responses for False Belief Understanding by Age Group (Study 3).**

<table>
<thead>
<tr>
<th>Response Classification</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Belief</td>
<td>66</td>
<td>72</td>
<td>140</td>
</tr>
<tr>
<td>Contamination</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Reality Error</td>
<td>33</td>
<td>12</td>
<td>45</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>104</strong></td>
<td><strong>88</strong></td>
<td><strong>194</strong></td>
</tr>
</tbody>
</table>

Individual subjects were then awarded a single point for each response classified as False Belief and given a False Belief score that ranged from 0–4 (see Table 14 below). Mean scores for each group were: Control= 3.00, SD=.90; False Belief=2.69, SD=1.35; Interpretive=2.94, SD=.93. No reliable differences in False Belief scores were found across conditions ($F[2, 45] = .376, p = .698$).

**Table 14: False Belief Understanding Scores by Age Group (Study 3).**

<table>
<thead>
<tr>
<th>False Belief Score</th>
<th>Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>2</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2</td>
</tr>
</tbody>
</table>

Subjects with False Belief Understanding scores of zero were classified as False Belief Absent, those with scores of 1 or 2 were termed Transitional, and scores of 3 and 4 were termed Present.
The results of this classification procedure are shown in Table 15. No reliable differences were found between groups in terms of the number of subjects in each category: \( \chi^2[4, N=48] =4.40, p=.35 \).

Table 15: Classification of Individual Subjects' False Belief Understanding by Age Group (Study 3).

<table>
<thead>
<tr>
<th>False Belief Understanding</th>
<th>Age Group</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td></td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Transitional</td>
<td></td>
<td>14</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td>10</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>26</td>
<td>22</td>
<td>48</td>
</tr>
</tbody>
</table>

With regard to false belief understanding, differences in performance were not observed between the training conditions in terms of raw scores, or classification of subjects. Age differences were apparent, however: the scores for 6-year-olds (\( M=3.27, SD=.767 \)) were reliably higher than those of 5-year-olds (\( M=2.54, SD=1.174, F[1, 46] =6.312, p<.02 \)). This echoes Taylor’s finding that 6-year-olds were more likely than 4-year-olds to correctly answer the question “Will [puppet’s name] know this is a ____?”. Despite these differences, all but 2 of the 48 subjects were able to attribute a false belief on at least one of the 4 occasions afforded them. While the training trials did not appear to differentially affect false belief performance across the conditions in this study, it is still possible that these manipulations could have affected performance across studies 2 and 3. That is, because these subjects were selected on the basis of their performance in Study 2, it is still possible that training could affect performance within subjects. In a Chi-Square analysis of subjects’ False Belief classifications in Study 2 and Study 3 (see Table 16) there is evidence of a significant migration of subjects from the Transitional classification on Study 2 to the Present classification on Study 3 (\( \chi^2[2, N=48] =23.05, p<.0001 \)). When the raw False Belief scores are used in a repeated measures ANOVA, no reliable effect for Training Condition (\( F[2, 45] =.231, p=.795 \)) was observed, however, a reliable effect for Study did emerge (\( F[1, 45] =161.3, p<.0001 \)) indicating higher False Belief scores in Study 3. The improvement in
false belief understanding, then, appears not to be connected with any of the particular training
conditions, but may instead be a function of further testing trials. Taylor reported a similar trend
for subjects in her Control Condition to perform somewhat better than did subjects in her first
study—a finding she attributes to the fact that these subjects experienced a “mild form of training”
(p.716) in hearing the puppet delivering incorrect responses.

Table 16: Contingency Table of False Belief Classifications (Study 2 and Study 3).

<table>
<thead>
<tr>
<th>Study 2</th>
<th>Absent</th>
<th>Transitional</th>
<th>Present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitional</td>
<td>2</td>
<td>17</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>Present</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>18</td>
<td>28</td>
<td>48</td>
</tr>
</tbody>
</table>

INTERPRETIVE UNDERSTANDING

As in Studies 1 & 2, the pair of responses attributed to the puppets for each stimulus picture
were coded for evidence of an understanding of interpretation. A pair was classified Interpretive if
both responses qualified as False Belief and were different from one another. All other response
pairs were classified as either Non-Interpretive–Repetition (if the same belief was attributed to both
puppets) or Non-Interpretive–Mixed (a combination of a False Belief and a Reality Error or
Contamination). Table 17 reports the results of these coding efforts by age group. By these
criteria, 6.7% of the response pairs given by 5-year-olds, and 8.0% of pairs from 6-year-olds
qualified as Interpretive (see Figure 12).

Table 17: Classification of Responses for Interpretive Understanding by Age Group (Study 3).

<table>
<thead>
<tr>
<th>Response Pair Classification</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretive</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Non-Interpretive–Mixed</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Non-Interpretive–Repetition</td>
<td>95</td>
<td>74</td>
<td>169</td>
</tr>
<tr>
<td>Total</td>
<td>104</td>
<td>88</td>
<td>192</td>
</tr>
</tbody>
</table>
Classifying individual subjects’ understanding of interpretation was accomplished by awarding them a single point for each Interpretive response pair they produced over the 4 trials. Scores could range from 0–4 (see Table 18). No differences were observed between the two age groups in terms of their Interpretive scores ($\chi^2 [3, N=48] =4.80, p=.187$).

**Table 18: Interpretive Understanding Scores by Age Group (Study 3).**

<table>
<thead>
<tr>
<th>Interpretive Score</th>
<th>Age Group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td></td>
<td>21</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6-year-olds</td>
<td></td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>40</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

These scores were then transformed to classify individual subjects by assigning the label Interpretive Understanding–Absent to those with scores of zero; Transitional to those with scores of 1 or 2; and Present to those with scores of 3 or 4 (see Table 19). Although the 6-year-olds performed somewhat better on this measure than did the 5-year-olds, the difference did not reach statistical significance ($\chi^2 [2, N=48] =4.464, p=.107$). Table 20 shows the Interpretive classifications by Condition. No reliable differences were found between the conditions ($\chi^2 [4, N=48] =2.35, p=.672$).
Overall, just 8 of the 48 (17%) subjects in Study 3 gave any indication of appreciating the interpretive aspects of the stimuli, compared with 46 of 48 (96%) who gave some evidence of understanding that such stimuli can promote false beliefs.

Within subjects comparisons were conducted on the Interpretive scores and classifications across Studies 2 and 3. Changes in Interpretive scores were recorded for 8 subjects (see Table 21). The scores of 4 subjects changed from zero to 1, two subjects scores increased from 1 to 2 and two subjects moved from 1 to 3. No score decreases were observed and a reliable difference between studies emerged: $\chi^2 [3, N=48] = 48.0, p<.0001$. A total of 6 subjects changed Interpretive classifications from Study 2 to Study 3 (see Table 22). Four subjects moved from the Absent to Transitional classifications and 2 from Transitional to Present—a sufficiently large number to result in a reliable difference between studies ($\chi^2 [2, N=48] = 30.54, p<.0001$). When the raw Interpretive scores are used in a repeated measures ANOVA, no reliable difference emerges for Condition ($F[2, 45] = 1.91, p=.313$), however, a reliable effect for Study did emerge ($F[1, 45] = 8.242, p<.01$) indicating higher overall Interpretive scores in Study 3. As with false belief understanding, it
would appear that if there is a training effect at work, it is a general effect and not specific to one type of instruction.

**Table 21: Changes in Subjects’ Interpretive Scores from Study 2 to Study 3.**

<table>
<thead>
<tr>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
</tr>
</tbody>
</table>

**Table 22: Changes in Subjects’ Interpretive Classifications from Study 2 to Study 3.**

<table>
<thead>
<tr>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>Absent</td>
</tr>
<tr>
<td>Transitional</td>
<td>Transitional</td>
</tr>
<tr>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
</tr>
</tbody>
</table>

**Comparing False Belief and Interpretive Understanding**

To test the overlap between measures of false belief and interpretive understanding, a contingency table was constructed using subjects’ False Belief and Interpretive classifications (see Table 23). A Chi-Square analysis of the reduced matrix (the non-shaded cells in Table 23) indicated that the measures were redundant ($\chi^2 [1, N=44] =1.69, p=.119$).

**Table 23: Contingency Table of Subjects’ False Belief and Interpretive Classifications (Study 3).**

<table>
<thead>
<tr>
<th>Understanding of Interpretation</th>
<th>Absent</th>
<th>Transitional</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transitional</td>
<td>17</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Present</td>
<td>21</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

It should be recalled that these subjects were specifically chosen from among Study 2 participants on the basis of their high False Belief and low Interpretive scores on these same measures, making
this analysis particularly sensitive to any movement out of the Interpretive Understanding–Absent category.

Discussion

The purpose of Study 3 was to determine whether or not Taylor’s (1988) method of presenting droodles actually taps subjects’ abilities to attribute false beliefs, and whether her training techniques can be adapted to influence subjects’ interpretive understanding. Subjects for the present study were chosen in a manner that would highlight any observed training effects. That is, subjects were selected on the basis of their performance in Study 2 such that all were known in advance to have at least some understanding of false belief, and to have little or no understanding of interpretation. Thus, if Taylor’s training techniques were to have any influence on what she terms “conceptual perspective taking” we could expect some improvement in the subjects’ understanding of either interpretation, false belief, or both. My own analysis of Taylor’s training methods held that these address only the possibility of false belief, leading to the prediction that subjects in the present False Belief Training Condition—which replicated Taylor’s procedure—might improve their false belief scores but not their interpretive scores. Similarly, it was anticipated that subjects in the present Interpretive Training Condition, in which the possibilities of multiple interpretation were modeled, might show improvement in interpretation but not false belief. The data, however, show a non-specific increase in both false belief and interpretive understanding across all conditions. Among the available interpretations of these findings are that either training of any kind is equally effective, or, more simply, that increased exposure to the testing materials and the procedures promotes better performance on the post-training outcome measure. With a single notable exception, the findings also come close matching the pattern of findings from Taylor’s Study 2. Although the test questions were different (open ended, vs. forced choice), and a new form of training was added, a general improvement in scores was noted in both studies, as was the tendency for all subjects to blame the puppet’s failures on the difficulty of the task rather than a lack of ability. The exception concerns the false belief performance of subjects in the False Belief training condition. While Taylor reported significant main effects for both age and condition
(the age by condition interaction was not significant)—findings that seem largely due to the near ceiling performance of the 6-year-olds—the present study revealed only a main effect for age, and again, the 6-year-olds’ performance approached ceiling levels (mean score=3.24 of a possible 4.0, compared with 2.54 for the 5-year-old subjects).

Whether a test-retest effect or not, the findings of the present study show, again, that there is a gap between false belief and interpretive understanding—that is, that when scored on both measures, subjects appear to find the attribution of a false belief to one puppet easier than attributing different false beliefs to two puppets. Along with the findings of Study 2, these data support the devotion of more concentrated attention on the question of why it is that some children who evidence a good grasp of false belief, by attributing a false belief to Ann, often fail to provide a second false belief for Andy. This question is taken up more directly in Studies 4–6.
**STUDY 4**

Studies 2 and 3 establish that children do not uniformly pass both the false belief and interpretive versions of the droodles task. The purpose of Study 4 was to examine some of the reasons for this uneven performance. One possibility is that children, even those who are interpretive, simply find it difficult or impossible to give more than one interpretation of a restricted view because the saliency of the underlying picture overwhelms their interpretive capabilities and all of their efforts are concentrated on the task of editing out their own privileged knowledge of the contents of the underlying picture in framing their understanding of the possible responses of others. The potential objection is that once a child knows that the underlying picture contains a ship and a witch, no amount of covering up bits of the picture will induce them to say anything other than “ship and witch.” There is no doubt that some measure of self-control is an unavoidable task demand within this testing context, however, we need to be first convinced that children can at least treat the droodle as a thing unto itself before going on to trust the procedure to reveal their interpretive capabilities. These concerns can be addressed by having Ann view the picture through one window and Andy view through another. To provide different false beliefs to the puppets would, in these circumstances, require only an appreciation of false belief. The simple question that this manipulation is meant to answer is: Can children see the restricted views as somehow things-in-themselves, distinct from the underlying picture, and then go on to differently label the separate windows?

It could also be the case that children fail to be interpretive only because they find no reason to assign different beliefs to the puppets. In other words, subjects otherwise capable of recognizing that two different target characters can sometimes entertain different interpretations of one and the same stimulus may simply see no compelling reason to assume that such differences are in place in this particular situation. Perhaps these subjects need some additional prodding to display their interpretive talents. The most blatant of helpful hints would be to outright tell the children that Ann and Andy disagree as to the contents of the restricted view. In Study 4, subjects were given just these instruction after they had provided an attribution for Ann and prior to being asked about
Andy’s thoughts. These changes will hopefully pave the way for them to report upon the puppets’ two different interpretations if they have it within them to do so.

Finally, it is possible that if subjects fail to provide a second interpretation of any droodle, they may do so not because they lack the conceptual wherewithal to recognize the possibility of interpretive diversity, but, more simply, because they themselves lack the imagination or creativity to actually invent a second reading of any one restricted view or, they may find these particular droodles to be too impoverished as stimuli to support a second reading. To help eliminate these possibilities, a final droodle was presented. In this last presentation, subjects were shown only a restricted view while Ann and Andy remain tucked away inside their houses. Subjects were asked to “guess what this [the restricted view] might be.” Following their initial response, the experimenter remarked: “That’s a good guess, now guess again. What do you think this is?” If any of the subjects were to find it difficult to have more than one thing to say about any one picture (note that the subject has not seen the full unrestricted view), their lack of imagination should become apparent in this final presentation.

Subjects

A total of 45 children (21 boys, 24 girls) were recruited for testing in this study. Fifteen children were tested in each of 3 age groups: 5-, 6-, and 7-year-olds. Mean ages for each group were: 5.52 years, 6.44 years, and 7.61 years, respectively. All subjects were recruited from, and tested in, a community-based after-school care facility in Vancouver, British Columbia. The experimenter spent several days in the after-school facility becoming familiar with the subjects prior to testing.

Materials

Materials consisted of the 7 droodles used in Study 1 (see Appendix III) and the puppets and boxes used in Studies 2 and 3. A small cassette recorder was used to record subjects’ responses for later transcription. Subjects were tested individually in a small office within their after-school care facility. The stimulus pictures used in this study were constructed using covers fitted with 2 non-
overlapping windows that allow two different restricted views to be produced from each picture (see Figure 13 below).

*Figure 13: Two window version of the ship-witch droodle.*

Design & Procedure

As in studies 2 and 3, subjects were introduced to Raggedy Ann and Raggedy Andy and asked to pretend that the dolls are “real people” and told that “when Ann and Andy are inside their houses, they can’t hear what we’re saying, and they can’t see what we’re doing.” The dolls were then placed inside their respective houses, and the subject was asked “Can Raggedy Ann hear us talking right now?”, and “Can Raggedy Ann see us right now?” A series of trials were then held that all conform to the following basic script:

The stimulus picture was presented with the cover opened to reveal the entire picture, and the child was asked: “What is this is a picture of?” If the child’s response failed to take into consideration those particular sections of the picture that would later be visible through the window
in the cover (for example, the bow of the ship or the witch’s hat), they were asked to specifically comment on those features of the picture.

The child was then told: “Now we’re going to get Andy out of his house and show him the picture. Andy has never seen this picture before. We’re going to show it to him like this...” at which point the cover of the picture was closed and one of the windows covering a restricted view was opened. The subject was then instructed to remove Andy from his ‘house’ and asked “What does Andy think that’s a picture of?” The subjects’ response was noted and repeated: “So, Andy thinks it’s a ______.” If the subjects’ attribution to Andy did not account for all of the details of the restricted view, they were asked to explain what Andy thinks “that part” was. Andy then remained at the subjects’ side for the duration of the trial.

The subject was then told: “Now we’re going to get Ann out of her house and show her the picture. Ann has never seen this picture before. We’re going to show it to her like this” and, again, a window onto a restricted view was opened. The subject was then instructed to retrieve Ann from her house and asked: “What does Ann think that’s a picture of?” Again, the subjects’ response was noted and repeated: “So, Ann thinks it’s a ______.” As before, if their attribution did not include some detail apparent in the restricted view, they were asked to explain what Ann thought of that detail. Ann and Andy were then returned to their ‘houses’ in preparation for the next trial.

A total of 6 droodles, divided into 3 groups of 2 droodles each, were presented. In Two Window Trials, both of the viewing windows were used. That is, Ann was shown a droodle through one viewing window, while Andy saw a different droodle through a second window on to the same underlying picture. The puppets therefore saw two different droodle views of the same underlying picture. Subjects were asked to describe what Ann would think of the contents of window 1, and then what Andy would think of the contents of window 2.

---

10 As in Studies 2 and 3, the presentation order for the puppets was counterbalanced across subjects rather than trials.
In *Single Window Trials*, a single viewing window was used, but this time, subjects were asked what first Ann and then Andy would say about the contents of one and the same window. Trials of this type constitute what will elsewhere be called “Standard Trials.”

*Disagree Trials* were, in most respects, a repeat of Single Window Trials in that inquiries were made about Ann and Andy’s reactions to the same droodle view. Here, however, the two characters were said by the experimenter to actively disagree about the contents of the restricted view. For example, if the subject said of the Witch droodle, that Ann would think it was “two knife points,” they were first reminded of that supposed fact and then told: “You know what? Andy doesn’t think it’s two knife points, what does Andy think it is?” This manipulation was designed as a hedge against the possibility that certain subjects might otherwise be quite capable of recognizing that two different target characters could sometimes entertain different interpretations of one and the same stimulus, but simply see no compelling reason to assume that such differences are in place in this particular situation.

In order to ensure that the subjects do not assume that Andy and Ann always, or automatically disagreed about the contents of the restricted views, the Disagree Trials were always conducted last. The order of the first two trial types was counterbalanced. As noted earlier, the order in which the dolls are removed from their houses, that is, whether Ann or Andy got first peek at any droodle, was held constant within subjects and counterbalanced across subjects. Finally, the droodles were presented in one of six possible orderings such that each droodle appeared in each trial type and serial position with equal frequency.

On the final trial using the 7th stimulus picture, subjects were shown only a restricted view while Ann and Andy remained inside their houses. Subjects were asked to “guess what this [the restricted view] might be.” Following their initial response, the experimenter remarked: “That’s a good guess, now guess again. What do you think this is?” The subject was then shown the full picture and told: “Those were good guesses! Here’s what the big picture looks like.”
Results

Subjects’ attributions to Ann and Andy on the various single window trials of this study (i.e., pairs of responses to the four droodles that make up Single Window and Disagree Trials) were scored using the criteria for false belief and interpretive understanding described in Study 1 (see page 49).

FALSE BELIEF UNDERSTANDING

All but two of the 45 children tested (two 5-year-olds) were able to attribute false beliefs to these puppet characters on at least one occasion. The percentage of all responses that could be classified as false belief attributions for the 5-, 6-, and 7-year-olds were: 53.3%, 68.3%, and 98% respectively.

Summary scores for each subject were derived by first assigning them one point for each response coded as a recognition of the possibility of false belief. Scores could thus range from 0-6. Using these scores, subjects who failed to evidence any understanding of false belief by responding, at every opportunity, with a reality error (i.e., those with summary scores of zero), were classified False Belief Understanding Absent. Subjects whose scores ranged from 1-3 were classified False Belief Understanding Transitional, and those whose scores ranged from 4-6 were classified False Belief Understanding Present. The distribution of these classifications by age group is shown in Table 24.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>False Belief Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absent</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>2</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>0</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
</tr>
</tbody>
</table>

As can be seen from an examination of this table, only 2 of the youngest subjects were unremittingly “realistic,” never attributing a false belief to Ann or Andy. Slightly more than a quarter of the subjects (12) evidenced some partial understanding of false belief, and over two-
thirds (31) appropriately attributed false beliefs to Andy and Ann at nearly every opportunity. As one might have anticipated, the group of subjects who were most successful in attributing false beliefs to the puppet figures, were, on average, older (6.7 years) than those in both the False Belief Understanding Absent (5.7 years) and Transitional (5.9 years) groups. No reliable differences were obtained between the groups in terms of gender ($\chi^2[1,45]=1.68, p=.43$), trial order ($\chi^2[5,45]=2.88, p=.24$), or puppet order ($\chi^2[1,45]=1.60, p=.45$). It should also be noted that, overall, more than 95% of the subjects correctly concluded, on at least one occasion, that these only partially informed puppet figures would end up holding to false beliefs about what they were looking at.

**INTERPRETIVE UNDERSTANDING**

A direct measure of when children of various ages first come to an interpretive or constructive theory of mind is provided by a comparison of subjects’ attributions to Ann and Andy on the various single window trials of this study. Clearly, subjects who understand what Flavell (1988) has characterized as the “one-many” relation that obtains between things in the world and their possible representations are in a position to recognize that while Andy might regard the droodle portion of a picture as being one thing (e.g., “kitty ears”), Ann might differently judge precisely the same ambiguous stimulus display as being something else entirely (e.g., “an arrowhead”). By contrast, subjects who have not yet arrived at such an interpretive view of the knowing process are essentially limited to either throwing up their hands, or simply repeating themselves by assuming that since both Ann and Andy are exposed to the same stimulus event, they are obliged to arrive at the same conclusion as to its identity. If, as was the case on Disagree Trials, where the puppets were said to disagree, such persistently realistic subjects are directly instructed to assign *different* interpretations to these objectively identical events, they should find it impossible to comply and end up by repeating themselves, by refusing to answer, or by falling back upon the realistic view that either Ann or Andy somehow knows the whole truth and need not make any interpretation at all.
In order to examine these several possibilities each subject’s responses to the four droodles that made up the Single Window and Disagree trials were coded using the criteria for interpretive understanding described in Study 1. Given a total of 60 opportunities (15 subjects giving 1 response pair for each of 4 droodles), the 5-year-old subjects managed to produce Interpretive response pairs on just 20% of these occasions. Repetitive pairs accounted for a full three-quarters of the responses, with Mixed pairs comprising 5%. For the 6-year-old subjects, an even third of the response pairs were classified as Interpretive, 57% as Repetitive, and 10% as Mixed. The 7-year-old subjects were much more consistent with 88% Interpretive, 12% Repetitive, and no Mixed responses. These results are summarized below in Figure 14.

*Figure 14: Interpretive Response Types by Age Group (Study 4).*

![Graph showing Interpretive, Repetitive, and Mixed response types by age group.]

Individual subjects were then assigned to one of three general scoring categories. Those with scores of 3 or 4 were labeled Interpretive Understanding—Present, those with scores of 1 or 2 were labeled Transitional and those with scores of zero were labeled Interpretive Understanding—Absent. By these criteria 40% of the subjects were found to be altogether unable to attribute different false beliefs to Ann and Andy, 13% managed to assign different false beliefs to the two doll figures on at least one occasion, and so were labeled as Transitional, and 47% qualified as having a well consolidated grasp of interpretation by attributing different false beliefs to Andy and Ann on nearly every available occasion (see Table 25). By these lights, just 60% of the subjects gave evidence of
possessing at least the beginnings of an interpretive theory of mind capable of generating two
different—yet both legitimate—false beliefs to the puppets who had viewed the self-same image.

Table 25: Interpretive Understanding by Age Group (Study 4).

<table>
<thead>
<tr>
<th>Interpretive Understanding</th>
<th>Absent</th>
<th>Transitional</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year-olds</td>
<td>10</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>8</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>0</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>6</td>
<td>21</td>
</tr>
</tbody>
</table>

Combining the *Transitional* and *Present* categories reveals a total of 27 subjects, or 60% of the
sample, who gave evidence of possessing at least the beginnings of an interpretive theory of mind
capable of generating two different—yet both legitimate—false beliefs to the puppets who had
viewed the self-same image. The attempt, on Disagree trials, to oblige subjects to provide different
responses by insisting that Ann and Andy disagreed, was largely unsuccessful. Only 5 of the 31
subjects who were not already at ceiling on Single Window trials actually improved their score on
Disagree trials, and all of these had already provided one interpretive response pair on Single
Window trials. No subject was prompted, by the manipulation on Disagree trials, to adopt an
interpretive view if they had not already done so on Single Window trials.

**CONTROL TRIAL**

It might have been that some of the subjects failed to provide a second interpretation of any
droodle simply because they lacked the imagination or creativity to actually invent a second reading
of the restricted view. To ward off this possible alternative interpretation, a final droodle was
presented and, after showing the subjects only the restricted view, they were asked “guess what
this [the restricted view] might be.” Following their initial response, the experimenter noted:
“That’s a good guess, now guess again. What do you think this is?” If any of the subjects were to
find it difficult to have more than one thing to say about any one picture, they should evidence their
lack of imagination here. In nearly all instances (97%), however, they provided guesses that
differed from one another in non-trivial ways. That is, when assessed against the criteria employed
in evaluating attributions to Andy and Ann, almost every response was judged *Interpretive*. What these control data suggest is that the subjects did not suffer any lack of imagination and that, when asked, such children can easily re-read the same stimulus picture in a new and different way.

**COMPARING FALSE BELIEF AND INTERPRETIVE UNDERSTANDING**

The final step in the analysis examined the extent to which measures of false belief and interpretive understanding are actually redundant. A contingency table was constructed that, based on their summary scores, boxed each of the subjects into one cell within the nine cell matrix shown in Table 26.

**Table 26:** *Contingency Table of Subjects’ False Belief Understanding and Interpretive Understanding (Study 4).*

<table>
<thead>
<tr>
<th>False Belief Understanding</th>
<th>Interpretive Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>Absent</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Transitional</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
</tr>
</tbody>
</table>

It should be noted that the scoring criteria for interpretive responses required that, in order to qualify as *Interpretive*, the first member of any pair of responses must have been scored *False Belief*. In other words, it was impossible for a response pair that did not begin with a false belief attribution to be scored as *Interpretive*. As a matter of definition, then all of the subjects classified as *Interpretive Understanding—Present* are also classified as *False Belief Understanding—Present*. Thus, only one cell in the column labeled *Interpretive Understanding—Present* could contain a number other than zero. Similarly, only the first cell in the row labeled *False Belief Understanding Absent* could contain a non-zero number. As in the earlier studies, in order to effect a legitimate comparison of false belief and interpretive theories of mind, it was judged prudent to remove the shaded cells from Table 26.

The fact that over 95% of the subjects were able to come up with at least one false belief, while only 60% of these same subjects could muster one or more interpretive response pairs, provides an
initial indication that having an understanding of the possibility of false belief is not the same thing as harbouring a constructivist theory of mind. A Chi-Square analysis of the non-shaded cells of Table 26 offers strong statistical confirmation of this view ($\chi^2(1, N=22) = 9.9$, p<.002, Fisher’s Exact p<.003).

Discussion

Study 4 represents the main test of the overall hypothesis that false belief understanding is not equivalent to an understanding of interpretation. Although the data support the view that the measures of false belief and interpretive understanding are not redundant and that children who pass one are not guaranteed to pass the other, the generality of these findings is limited in several ways. First, the data do not rule out the possibility that children may be confused by the stimuli used in these studies. These are, after all, small, deliberately cryptic, 2-dimensional line drawings that may not lend themselves to alternative interpretations (or may not do so in the eyes of children rather than adults as was the case in Study 1). What is still needed, then, is a demonstration that children of this age can, in fact, see multiple possibilities in these particular stimuli. That task is undertaken in Study 5.

Another limitation of this study lies in the fact that subjects may be compromised in their ability to arrive at a second interpretation because the reality of the picture continues to stare them in the face. Taylor (1988) suggested that perhaps her subjects were much less likely to attribute action and personal knowledge (the giraffe is sitting, his name is George) to the puppet than they were to attribute knowledge of the objects’ identity because part of the information needed to make a proper identification is always visible. It may be then, that the extra effort required to first cordon off what one knows about the larger picture before going on to invent an alternative reading is sufficient to upset this latter process. In Study 6, measures are taken to overcome this potential limitation by attempting to provide the subjects with “ready-made” interpretations that could be legitimately attributed to the puppets.
STUDY 5

It could be argued that, despite the inclusion of a seventh droole in Study 4, one cannot be sure that the failure of some children to give interpretive responses is not still somehow an artifact of the stimuli used rather than a conceptual difficulty on the part of the subjects. What we really need to know, then, is whether or not children can easily generate alternative interpretations, not of cryptic line drawn pictures in general (as the data from the 7th droole shows), but of the six particular drooolles actually used in Study 4. This rather straightforward question can be answered using a new sample of subjects and presenting them with these 6 drooolles using the procedure described for the 7th droole in Study 4. That is, if subjects asked to “guess” and “guess again” can give two different readings, we can be reasonably sure that the participants in Study 4 were able (in principle) to similarly generate alternate, non-interpretive, readings of these stimuli.

Subjects

A total of 30 children (13 boys, 17 girls) were tested in this study (10 each of 5-, 6-, and 7-year-olds). Mean ages for each group were: 5.41 years, 6.58 years, and 7.46 years. All subjects were recruited from, and tested in, a community-based after-school care facility in Vancouver, British Columbia.

Materials

Materials consisted of the 6 drooolles and other experimental materials used in Study 4 (see Appendix III). A small cassette recorder was used to record subjects’ responses for later transcription.

Design & Procedure

Subjects were tested individually in a small office within their after-school care facility. Subjects were shown a restricted view of one of the stimulus pictures and asked to: “Guess what this is.” The experimenter then noted: “That’s a good guess, now guess again. What do you think this is?”
Each subject was tested on 3 of the 6 droodles used in Study 4, with droodles counterbalanced across subjects.

Results

Subjects' responses to the “guess what this is” question were scored using the same criteria employed in Study 1. The question of interest concerns the frequency with which subjects gave two different “guesses” in response to each droodle. The short answer is: almost always. The 6- and 7-year-old subjects performed flawlessly with each and every child giving two different “guesses” to each and every droodle they were shown. Examples of responses given to the Witch droodle include: “two knives,” “two pointy sticks,” “an arrow,” and “sharp teeth”—essentially the same sorts of responses that were offered by subjects in Studies 1 and 4 as possible beliefs that Ann or Andy might entertain. Nine of the ten 5-year-olds gave perfect performances, with just one child failing to give a second guess on two of the three occasions she was afforded. The percentage of correct responses (a correct response would be giving a pair of different guesses concerning any one droodle) by age group are: 5-year-olds 93%; 6-year-olds 100%; 7-year-olds 100%. This level of performance rendered any further search for particularly “easy” or “hard” stimuli, or for age or gender differences, unnecessary.

Discussion

These findings clearly support the data from Study 4 suggesting that subjects in this age range do not suffer any lack of imagination and, when asked, such children can easily re-read the same stimulus picture in a new and different way. More importantly, this study effectively rules out the possibility that the subjects in Study 4 who failed to offer interpretive responses to these same stimuli might have done so because the stimuli are somehow inherently difficult to re-read.

The conclusion to be drawn from this study seems to be that, whatever else might serve to bar the way of children’s easy progress in coming to an interpretive view of mind, we can at least eliminate the ability to say more than one thing about a particular picture, or to imaginatively re-read the same picture in a new and different way. While such a skill is obviously required to pass
the task set out in Study 4, it is seems unlikely that the failing performance of some children on that task could have been the result of an inability to re-read the droodles presented. As noted earlier, however, the subjects in Study 4 may have experienced difficulty in attributing a different belief to the second puppet for other reasons—reasons that are examined in Study 6.
The findings of Studies 2–5 might be challenged from several perspectives. According to one line of reasoning, the failure of some subjects to attribute different false beliefs to Ann and Andy (i.e., to provide *Interpretive* response pairs) might arise not from a simple lack of imagination (which Study 5 rules out), or a pervasive failure to appreciate that beliefs about the picture must be inferred from the ambiguous stimulus (most subjects clearly already appreciate this, as evidenced by their understanding of false belief), but rather because their ability to generate plausible alternative beliefs, which may be fragile at the best of times, cannot match the task demands of this particular assessment situation. That is, much like adults made to view a very large window onto a droodle—a window so large that it fails to obscure any meaningful portion of the picture beneath—the young subjects in Study 4 (but not those in Study 5) may have been unable to generate or use plausible alternative interpretations because their knowledge of what the picture actually depicts so overwhelms their ability to entertain other possibilities, that they are limited either to repeating themselves, to committing a reality error, or to falling silent.

This reasoning would imply that performance would improve if one could somehow provide the young subjects with some ready alternative interpretations. One could, for example, add a pre-test trial, similar to the procedure followed in Study 5, in which subjects had access to alternative interpretations of precisely the droodle that was about to become the focus of attention by first showing the subject the restricted view and asking: “What do you think that is?” followed by: “That’s a good guess, have another guess.” Since the child has not yet seen the full picture, these guesses (provided they are incorrect) become, when the picture is revealed, false beliefs about its contents. On this account, when faced with the task of attributing a plausible belief to Ann and then to Andy, children with even the most tenuous grasp on the interpretive nature of the knowing process could reasonably be expected to take advantage of their own previous ‘guesses’ and to attribute these as legitimate false beliefs to the puppets. Of course, we would need to be convinced that these alternatives were actually available to the subject ‘on-line’ or during the trial itself. Still, a
simple post-test recall question ("What did you guess when I first showed you the picture?") could ensure that the alternative representations weren't somehow flushed from the subjects' memory when their access to the full picture had proven their guesses "wrong." In Study 6, a sample of thirty 5-to 7-year-old subjects was tested using these modifications to the earlier procedures.

At issue in this study was whether or not having subjects first guess at the contents of the droodle would serve to provide them with plausible alternative interpretations that could later be attributed to Andy and Ann. The data would be probed to determine: (a) whether or not subjects' performance on the task in terms of false belief and interpretive understanding would be enhanced by this manipulation; (b) the extent to which their own alternative interpretations (i.e., guesses) would actually be used as attributions to Andy or Ann as false beliefs; and (c) whether, if these guesses were not utilized, we could be sure that they had been available during the trial (i.e., could the subjects later recall their guesses).

Finally, in addition to shoring up the findings from earlier studies, this last study was designed to provide a more detailed understanding of how the subjects themselves view the task of attributing beliefs to the puppets. Toward this end, the most competent and articulate group of the subjects (the 7-year-olds) were asked a series of questions meant to prod them into justifying the beliefs they attribute to Andy and Ann. Following each of the final 2 trials, the 7-year-old subjects were asked the following questions:

1. Why would Ann think it was a [belief attributed by child]?
2. Why would Andy think it was a [belief attributed by child]?
3. Why didn’t they think the same thing?
4. One child said Andy and Ann would think the same thing, is that OK?
5. Would Ann or Andy know it was a [description of full view] from just seeing this [restricted view]?

It was hoped that these questions would not only lead the subjects to discuss the interpretive possibilities afforded by the testing procedure, but also to provide some insight into the process by which their attributions were constructed.
Subjects

A total of 30 children (16 boys, 14 girls) were tested in this study (10 each of 5-, 6-, and 7-year-olds). Mean ages for each group were: 5.61 years, 6.43 years, and 7.29 years. All subjects were recruited from, and tested in, a community-based after-school care facility in Vancouver, British Columbia.

Materials

Materials consisted of the 6 droodles used in Study 1 (see Appendix III). A small cassette recorder was used to record subjects’ responses for later transcription.

Design & Procedure

Subjects were tested individually in a small office within their after-school care facility. Six droodles comprising 3 trial types of two droodles each were presented, but here only one-window trials were used—that is, Andy and Ann always viewed the same restricted window onto any droodle.

Trials 1 and 2 formed the Standard Trials in which subjects were asked what first Ann and then Andy would “think” of the restricted view. Trials 3 and 4 formed the Guess Trials, which was in all other respects identical to the Standard trials with the exception that subjects were first shown the restricted view and asked to “Guess what this is.” Following one guess, they were told “That’s a good guess, guess again.” Once two guesses had been made, subjects were shown the unrestricted view before being asked about Andy and Ann’s beliefs as in the Standard Trials. Trials 5 and 6 formed the Guess+Recall Trials. This trial pair was identical to those in the Guess Trials except that, at the end of each trial, subjects were asked “Remember when I first showed this to you this [restricted view], what was your first guess? What was your second guess?” Following trials 5 and 6 the subjects were asked the post-test questions listed above.
Results

FALSE BELIEF UNDERSTANDING

When submitted to the scoring and coding criteria employed in Study 1, a pattern of findings similar to that of earlier studies emerged. Among the 5-year-old subjects, all were able to attribute at least one false belief, with 48% of all responses classified as expressions of false belief. Among the 6-year-olds, 78% of responses were classified as false beliefs, with all but one subject providing at least one false belief response. All of the 7-year-old subjects provided at least one false belief, with 92% of their responses classified as false belief. Using the same classification criteria employed in Study 1 resulted in the distribution of False Belief Understanding shown in Table 27.

Table 27: False Belief Understanding by Age Group (Study 6).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Absent</th>
<th>Transitional</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>0</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>6-year olds</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>9</td>
<td>20</td>
</tr>
</tbody>
</table>

As can be seen from an examination of Table 27, only one subject consistently exhibited errors of intellectual realism and never attributed a false belief, about one third of the subjects (9) correctly made at least one such false belief attribution, and a full two-thirds of the subjects (20) attributed false beliefs at every opportunity. As was the case in previous studies, the False Belief Understanding—Present subjects were, on average, older (6.9 years) than those in the Transitional group (5.7 years), and the single False Belief: Absent subject (6.5 years). Still, 97% of these subjects were able to attribute a false belief on at least one occasion.

INTERPRETIVE UNDERSTANDING

Using the same criteria for Interpretive, Repetitive, and Mixed response pairs employed in the earlier studies, the 5-year-old subjects in Study 6 managed to produce Interpretive response pairs
on just 12% of the possible occasions afforded them. Nearly three-quarters of their response pairs (73%) were coded as Repetitive, and 15% as Mixed. The 6-year-olds produced 45% Interpretive, 43% Repetitive, and 12% Mixed response pairs. The 7-year-olds showed 78% Interpretive, 15% Repetitive, and 7% Mixed responses.

Applying the criteria used in Study 1 to classify individual subjects, 30% of the subjects were classified as Interpretive Understanding–Absent, 30% as Transitional, and 40% as Present. The results by age group are shown in Table 28.

Table 28: Interpretive Understanding by Age Group (Study 6).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Non-Interpretive</th>
<th>Transitional</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>0</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

When, as before, the matrices of False Belief and Interpretive Understanding are collapsed into a single contingency table, subjects in Study 6 align themselves as shown in Table 29 below.

Table 29: Contingency Table of Subjects’ False Belief and Interpretive Understanding (Study 6).

<table>
<thead>
<tr>
<th>Understanding of False Belief</th>
<th>Non-interpretive</th>
<th>Transitional</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Transitional</td>
<td>7</td>
<td>3</td>
<td>--</td>
</tr>
<tr>
<td>Present</td>
<td>1</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>

While all but one of the 30 subjects could attribute a false belief to Ann, only 21 of these same subjects attributed a different false belief to Andy. As before, when the cells in Table 29 that should be vacant by definition are eliminated, along with those housing subjects whose performance is at ceiling or floor (i.e., the shaded cells of row 1 and column 3), a statistically
reliable deviation from the chance placement of subjects was found ($\chi^2[1, N=17] = 5.13, p < .03, \text{Fisher's Exact } p < .05$).

Finally, in comparison with the results of Study 4, the guessing manipulation did not appear to enhance false belief performance (67% of subjects classified as \textit{False Belief Understanding-Present} in Study 6 vs. 69% in Study 4), nor result in reliably higher scores on measures of interpretive understanding (40% classified as \textit{Interpretive Understanding-Present} in Study 6 vs. 46% in Study 4).

**USE & RECALL OF GUESSES**

When asked to provide two guesses in response to viewing the droodles, subjects in Study 6 performed in much the same fashion as did subjects in Study 5: subjects were able to provide two non-trivially different guesses as to the contents of the window on 114 of the 120 (95%) opportunities afforded them. The remaining 6 pairs included four repetitions of the first guess when asked for a second, and two failures—both on the part of a single child—to provide a second guess. During testing, subjects attributed these guesses as beliefs to Ann or Andy on just 39 of 240 possible occasions (16%) with nearly a third of these coming from just two subjects (one 5-year-old and one 6-year-old). Finally, subjects were able to correctly recall their guesses at the conclusion of 118 of the 120 trials (98%).

What these data suggest is that: (a) children find the production and recall of such guesses a simple matter, and; (b) that only rarely do their guesses end up coinciding with the beliefs they attribute to the puppets.

**JUSTIFICATION OF BELIEFS**

When asked to justify or explain their attributions, all of the 7-year-old subjects were quick to point out features of the droodle that might lead one to hold just the beliefs they had ascribed to the puppets. These off-the-cuff reconstructions of how Ann and Andy had been ‘led astray’ were typically accompanied by gestures and narratives concerning the orientation and composition of the
various ‘parts’ of the imagined picture and routinely included allusions to those parts which could be imagined to extend beyond the borders that framed the restricted view. Indeed, no subject was struck dumb by this question, and most had to be restrained from continuing their elaborations upon the puppets’ supposed beliefs.

Having given what, in their own minds, must have seemed quite adequate explanations, most subjects simply scoffed when asked why Ann and Andy had not come to the same interpretation of the droodle. When pressed, however, only two especially disappointed subjects retreated and claimed not to know. Most were able to revive their narratives and supply some sort of cogent answer. One child noted that, in her experience, Ann and Andy would “never agree on anything.” Another felt that Andy’s mischievous nature had compelled him to listen to Ann’s response and then “go her one better” by inventing a more fantastic interpretation. Four of the 10 subjects made reference to the context and the ambiguity of the stimulus, claiming that there was no fault to be found in Andy and Ann’s disagreement, since “that’s just how they saw” the droodle at that particular moment, and that their beliefs could easily have been otherwise, or even reversed. Two subjects sought to resolve the differences between Andy and Ann’s interpretations by suggesting that although we might not know what had led them to their differing views, there are always reasons for such differences and that more detailed information about Andy and Ann would be required to determine what those reasons were. Finally, one particularly savvy subject noted: “They’re not a two-headed monster with the same brain! Do you think they both think the same thing all the time?”

Fully half of the subjects balked when told that “one child said Andy and Ann would think the same thing.” The flavour of these responses is best captured by one of the subjects who rolled her eyes and muttered “yeah, right.” A further three subjects, though incredulous, agreed that while such a thing could conceivably occur, it was extremely unlikely. Though the remaining two subjects acknowledged that it was indeed “OK” to think such things, it was apparent that they meant “permissible” rather than “plausible.” Finally, all of these 7-year-old subjects emphatically
denied that it would be possible for Andy or Ann to know what the contents of the larger picture would be given access only to the restricted view.

The children gave good evidence to the effect that they understand: (1) that no one could be expected to know what the true contents of the pictures were, and; (2) that the beliefs they attribute to the puppets should be justifiable rather than simply flights of fancy, and; (3) that it would be perhaps permissible, but still particularly bad form, to attribute the same beliefs to both puppets.

Discussion
The findings of Chandler and Helm (1984) suggest that subjects' performance improved after “having stood precisely in another's shoes” by being cast in the role of the other and asked what they themselves think about a particular droodle. As noted earlier, the way in which the trials were conducted in Study 6 may have approximated this effect of “shared experience.” The guessing manipulation employed in the present study did not, however, dramatically enhance performance.

It seems likely that subjects experienced the guessing and belief attribution phases of the trials very differently. That is, when the subject had not seen the full picture and the real truth of the matter was still unknown, being asked to “guess what this is” involved generating a set of responses that were not seen by the subject to be statements about reality, but rather off-the-cuff propositions that no one could be expected to be held to account over. That children seem to grasp this aspect of the task can be inferred from the rising inflection of their responses (e.g., “A cow?”), indicating that they expect corrective feedback. Once the contents of the picture are known, however, and attention is turned to the thoughts of others, a very different context is created. Now the child must keep reality at bay and make use only of the visible parts of the droodle to generate a credible belief to attribute to the puppet. The data already in hand from the previous studies, tell us that children of this age can imaginatively re-read the stimuli when asked to guess. The data also show that such children can attribute false beliefs in this situation. Some subjects (especially the youngest) tend to repeat their false beliefs when asked about a second puppet, and do so even when told that the puppet will not share that belief. Because changes to the earlier procedures incorporated in the
present effort did not enhance performance, it seems even more likely that there is a very real
distinction between false belief understanding and interpretation.

The post-test interviews conducted with the 7-year-old subjects help reinforce the contention
first raised in Study 1, that droodles are seen by these children as objects that require interpretation.
Their responses revealed an understanding of the basic premises of interpretation: that the beliefs
must be justifiable, that multiple beliefs are all but inevitable in this context, and that it is somehow
a sign of immaturity to attribute full knowledge of the picture or exactly the same false belief to
both puppets.

This final study brings together many of the elements of the studies that preceded it. It offered
another opportunity to compare subjects' ability to see the possibilities for both false belief and
interpretation that are inherent in the ambiguous droodle stimuli. Again, steps were taken to bolster
the performance of those who might otherwise be held back only by extraneous demands of the
testing context. Having subjects first take the perspective of the puppets by being asked to
themselves guess what the droodle might be should have armed them with a pair of ready-made,
legitimate false beliefs about the picture. At the very least, this guessing manipulation should have
eased the cognitive burden of trying to come up with words to place in the mouths of the puppets.
Because these last attempts to narrow the gap between false belief and interpretive capabilities also
largely failed, yet more evidence was amassed in support of the main hypothesis that false belief
understanding is a necessary but insufficient condition for an appreciation of interpretation.
The overall aim of the sequence of studies just reported was to investigate young children’s earliest insights into the interpretive nature of knowing. It was argued that, at a minimum, such insight would need to encompass an understanding that it is possible for two persons to be exposed to precisely the same information or stimulus event and yet to arrive at different opinions about what is still the self-same reality. To understand mental life in this way would mark one as having acquired a theory of mind that shared in the common adult conception that minds not only capture and record information about an external reality, but also work in the other direction by creatively constructing and uniquely interpreting reality. This understanding of the interpretive character of the mind, it was argued, involves more than an appreciation of the possibility of false belief, and in opposition to certain readings of earlier research findings, the hypothesis tested in the present work held that false belief understanding is a necessary but not sufficient condition for this more constructive view of the knowing process. What remains to be accomplished in this concluding section is to first review the findings of the six studies reported above and then to discuss both the limitations and possibilities for further research that these findings bring out.

By way of review, then, the study sequence began with what amounted to a background task. In Study 1, the suitability of the stimuli and procedures were tested using adult subjects who could be safely assumed to already hold to an interpretive view. The goal of this first study was to ensure that the experimental stimuli were actually—as advertised—capable of prompting both false beliefs and multiple interpretations. The results indicated that both the subjects and the stimuli were equal to the task. Adults provided false beliefs on over 91% of all occasions afforded them, and then went on to provide differing interpretations of the same stimulus picture on 70% of all trials. Similarly, the individual stimulus pictures were each able to garner a sufficient number of unique interpretations and to avoid particularly common or popular readings to effectively satisfy any concerns about their suitability as objects of interpretation. With these results in hand, attention was then turned, in Study 2, to the reactions of young children to these same stimuli and procedures.
The results of Study 2 suggest that children of 5 years of age and older can, for the most part, be counted on to recognize the potential that these stimuli hold for promoting false beliefs. Only a subset of these same children, however, proved capable of appreciating the interpretive possibilities inherent in the pictures. What was also made clear in the findings of this study was that in order to properly assess children's understanding of both false belief and interpretation, one must first provide the subject with a full understanding of the real truth of the matter. In other words, it was only when the subject knew the real truth, while the puppets did not, that one could adequately assess whether or not they understood: (1) that the puppets might get things wrong (false belief), and/or; (2) that they might get things differently wrong (interpretation).

In an effort to improve the performance of a specially chosen subset of these children who had, in Study 2, shown some evident grasp of false belief, but little in the way of interpretive understanding, efforts were made in Study 3 to expose the same children to various training schemes. The study partially replicated procedures used by Taylor (1988) meant to showcase the ambiguity of the restricted views, with the addition of a new condition that modeled differing interpretations of the same stimulus pictures. While subjects' performance did not differ across these training conditions, an overall improvement was observed from their earlier performance in Study 2. While these findings may indicate that a test-retest effect is at work, it should also be noted that, as in Study 2, although all of these subjects could attribute false beliefs to the puppets, only a handful could ascribe differing interpretations of the same stimulus picture.

In Study 4, several alternative explanations for the now evident gap in performance on measures of false belief and interpretive understanding were addressed. First, it was shown that children are easily able to provide differing false beliefs when the puppets were made to view the same picture through different windows. That is, when Ann saw one restricted view while Andy saw another, subjects had little trouble attributing different beliefs. Second, this study ruled out the possibility that the children were generally capable of attributing different beliefs to the puppets but simply saw little reason to exercise that capacity in this particular testing environment. The results of Study 4 also indicate that even explicitly telling children that the puppets disagree about the
contents of the restricted view does little to improve their performance. Specifically, children who had not already attributed different beliefs to the puppets were not prompted to suddenly do so by this procedural manipulation. Finally, a short post-test measure showed that these children did not suffer some lack of imagination that prevented them from re-reading a restricted view in more than one way. What these results demonstrate is that children do not lack the required creativity or somehow misunderstand the task in ways that prevent them from exercising an appreciation of interpretation if they have it within themselves to do so. Finally, and once again, only a subset of those who passed the measure of false belief went on to pass the measure of interpretation.

In Study 5 yet another attempt was made to rule out reductive readings of the earlier findings by ensuring that the stimuli used in Study 4 were not somehow particularly hard for children to re-read. Here children were simply asked (without ever seeing the full pictures) to provide two guesses as to the likely contents of each of the restricted views used in Study 4. Almost without exception, the children were able to provide what, in other circumstances, would amount to “false” beliefs and “interpretive” response pairs for each and every restricted view placed in front of them. It would appear, then, that the failure of some children to behave interpretively in earlier studies is not the result of some artifact having to do with these particular stimuli.

Study 6 had two objectives. First, a final attempt was mounted to close the performance gap between the measures by providing children with ready-made alternative interpretations, and second, a post-test interview was conducted that aimed to probe the thoughts of the most capable of the subjects (the 7-year-olds) concerning the actual process of attributing beliefs to the puppets. The attempt to improve subjects’ interpretive scores largely failed. Despite having themselves given what amounted to different “false” beliefs about the same restricted view when asked to guess at its contents, very few subjects took advantage of the opportunity to employ these alternative readings as the beliefs they attributed to the puppets. In the post-test interview, the children described their experience in much the same way a typical adult might have done. That is, when asked why it was that the puppets had different ideas about the contents of the pictures, these children were quick to point to the fact that, since no one could possibly know the true contents, and since many different
things could potentially be said about the restricted view, it was highly unlikely (in their opinion) that any two persons would see and interpret these ambiguous lines in exactly the same way.

There is, perhaps, no better way to summarize the results of all six studies than to return to the data and note that a total of 189 children were given a total of 866 opportunities to demonstrate their understanding of false belief and 716 opportunities to demonstrate an understanding of interpretation. The children succeeded in attributing a false belief to Ann on 73.4% of these occasions, while on only 35.6% of these occasions did the same children go on to evidence an appreciation of interpretation by attributing a different false belief to Andy. This same pattern of findings appears when a contingency table is constructed using the False Belief and Interpretive classifications derived from Studies 2, 3, 4, and 6 (see Table 30, below). Overall, 116 of the subjects (61.4%) had a solid grasp of the possibility of false belief, while just 50 (26.4%) appreciated the possibility of interpretation. A Chi Square analysis of the non-shaded cells in Table 30 indicates that the measures are not redundant and (once again) that false belief understanding is not the same thing as an understanding of interpretation ($\chi^2[1, N=132]=12.261, p<.0005$, Fisher’s Exact $p<.0008$).

Table 30: Contingency Table of Subjects’ False Belief and Interpretive Understanding (Studies 2, 3, 4, and 6).

<table>
<thead>
<tr>
<th>False Belief Understanding</th>
<th>Interpretive Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-interpretive</td>
<td>Transitional</td>
</tr>
<tr>
<td>Absent</td>
<td>7</td>
</tr>
<tr>
<td>Transitional</td>
<td>61</td>
</tr>
<tr>
<td>Present</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>113</td>
</tr>
</tbody>
</table>

Despite the good intentions and careful design of these studies, however, a short list of reasons as to why the reader might remain unconvinced of the claim that false belief understanding is somehow different from a grasp of the interpretive nature of knowing would need to include at least the following three items. First, because droodles (especially Price’s originals) are
intentionally designed to be so devilishly cryptic, children might view them as being so bereft of meaning as to legitimate any sort of simple flight of fancy or verbalized day-dream one might choose. That is, rather than constituting objects requiring warrantable interpretation, the droolles stimuli might somehow work to gift child subjects with an undeserved mask of competence by falsely suggesting that they possess a genuine grasp of the concept of interpretation when, in fact, they do not. But the opposite possibility also exists: The fanciful nature of these tasks might have failed to fully engage the true ability of young persons to appreciate the inherently subjective nature of knowing. Although it could be argued that these potential false negatives and false positives effectively cancel one another out, both prospects raise the possibility that the droolle procedure frankly lacks a true epistemic component. What seems sorely needed, then, to bolster the findings of the present assessment efforts are other future studies that examine more common or canonical situations in which children’s early interpretive capabilities might be more naturally called into play.

The search for some better, more naturalistic testing situation need not take place outside of the assumptions that framed the current studies, however. The task that I employed was meant to allow child subjects to give some clear evidence that they understood that more than one defensible interpretation of a single object or event is possible by presenting them with stimuli that afforded many possible interpretations. The alternate procedure employed by Carpendale and Chandler (Carpendale, 1995; Carpendale & Chandler, in press) took a different tack by using stimuli drawn from the opposite end of this same dimension—stimuli that admit two and just two mutually exclusive possibilities. The rationale for sampling stimuli from the ends of this continuum was that children might find it easiest to reason about interpretive matters in instances in which the number of possible interpretations are so restricted as to rule out all but the minimum number needed to qualify as an instance of interpretation (i.e., two), or so numerous as to rule out almost no prospective candidate response. The results of these efforts (i.e., Carpendale’s and my own) show that by roughly 6 or 7 years of age, children are beginning to appreciate the interpretive possibilities inherent in stimuli of the foregoing sorts.
As noted above, several objections to the meaning of these findings could be raised. The most serious of these is that performance on the tasks provides a window onto a competence that is irregularly manifested or, worse still, that the procedures construct so efficient a scaffold as to conjure up success in the absence of any real underlying competence. Like a form of experimental entrapment, such tasks might function to create testing situations that are so well engineered that they mistakenly appear to call out interpretive capabilities that aren’t really there. In defense of both of these procedures, it must be noted that the subjects’ post hoc explanations and justifications give us some measure of assurance that they truly are grappling with real interpretive matters during the assessment. Still, visual illusions and droodles, which either beam out their two interpretations, or remain amenable to practically any old interpretation one might care to give, may not constitute the most appropriate of test objects. What is obviously required, in addition to, rather than in the place of, such measurement opportunities, are others that keep further back from the edge of these extremes.

While the possibilities for further research are wide indeed, two especially promising options present themselves. The first would be to follow the lead of researchers who have argued that the standard set of false belief measures are unnecessarily dehumanized and decontextualized and who have attempted to introduce some degree of realism and subject involvement to those earlier methods (Chandler et al., 1989; Fritz, 1992; Hala, 1994; Lewis, 1994). In the droodles procedure, for example, subjects could be allowed to determine the size and position of the window onto the larger picture, or asked to find windows that are particularly effective at promoting misinterpretations. Another approach would be to add an interpersonal dimension to the procedure by examining children’s understanding of those potentially embarrassing “lipstick on your collar” situations in which one’s innocent actions can be easily and distressingly misinterpreted. Reactions to these compromising situations would go some distance toward advancing our understanding of the how an appreciation of interpretation actually operates in the social and interpersonal lives of young persons.
Filling in the still missing details of our understanding of the role that an interpretive stance plays in the lives of young school-aged children is, however, not at all the same thing as having a developmental story to tell. That is, updating the “one-miracle” view of children’s developing theories of mind—a view that is already being abandoned by some of its proponents—by simply adding a second miracle may be a necessary first step, but it hardly constitutes real progress.

On a list of topics that offer some hope of revealing more about children’s growing appreciation of the subtleties of mental life, I count three especially promising candidates. Taken chronologically, these include: (1) the gradual elaboration of personological variables in a developing theory of mind—particularly as seen in the capacity of middle school children to incorporate notions of enduring personality structures in their explanations and predictions of social and interpersonal behaviour; (2) the movement, in adolescence, from entertaining retail doubts about the validity of some knowledge claims, to a wholesale rejection of the adequacy of any and every claim to certainty, and; (3) the intellectual struggles that plague essentially all of one’s tenure as a mature adult to justify and uphold a personally coherent world view and sense of personal identity in the face of challenge and change. Although each of these broad topic areas obviously deserve much more space than can be allotted them here, a few words will need to said if only to support their candidacy as items on a theories of mind “to-do” list.

With regard to the first topic, what might be called “child as personality theorist,” there is already good evidence to suggest that between the ages of 6 and 11 years children move from giving situational or behavioural accounts of the behaviour of others to more sophisticated psychological descriptions and comparisons that include the notion of stable personality traits (Barenboim, 1981; Rholes & Ruble, 1984). This seems fertile ground indeed for a researcher interested in the changing ways that children come to make real use of these powerful but more intricate variables in their theorizing about the mental lives of others and in making concrete judgments about just how it is that a particular peer or playmate might interpret the nuances of a given situation.
Slightly farther along the developmental course are questions concerning the ways in which somewhat older children and adolescents resolve matters of intellectual uncertainty. That is, how such young persons come to first recognize, and later to overcome an awareness of their own fundamental “epistemological loneliness” (Chandler, 1975) or ultimate subjectivity in a world marked by the unrelentingly private and privileged nature of their own and others’ thoughts and feelings. This disquieting realization foreshadows even more serious epistemological troubles: “unlike their younger counterparts, such adolescent youths no longer hold out the prospect that all differences of opinion will eventually succumb to the authority of objective facts, and instead toy with the prospect that certain of the interpretive differences that divide people are endemic to the knowing process itself” (Chandler, 1988, p.409). Coming to grips with the essential subjectivity of the knowing process and the notion that all knowledge is unavoidably interpretive, brings the adolescent squarely to the problem of the fundamental ambiguity of all knowledge. Here, where all claims to absolute certainty must be abandoned, they are left to construct a theory of mind that rests on the impossibility of maintaining a “viable, certainty-preserving bridge between the realms of subjective experience and objective truth” (Chandler, 1988, p.411).

The prospects for making individual headway on such a seemingly bleak epistemological landscape are not entirely foreclosed, however, and a substantial body of more hopeful research exists concerning the routine business of making one’s grown-up way through a world of ultimately undecideable knowledge claims. Among the ranks of those charting progress toward the upper limits of an interpretive theory of mind, the work of Perry (1970), Kitchener and King (1981) and those researchers concerned with women’s epistemic development (Belenky, Clinchy, Goldberger & Tarule, 1988) deserves special mention.

The connection that is offered here between the findings of my own studies of young school-aged children and those arising out of the ongoing work of others examining later developmental periods, is rather simple: Children begin their careers as theorists of the mind by recognizing that others can ‘get things wrong’ (simple false belief); they progress to the realization that there is more than one way to be wrong (interpretation); then to more sophisticated notions of interpretation
that include the possibility that there is (perhaps) more than one way to be ‘right’; and eventually to
the view that any and all claims to knowledge rest upon finding the best available warrant for what
must remain fundamentally doubtable knowledge claims. Updating the “one-miracle” view to
encompass all of these findings is not so simple and, thankfully, is not a task that need be
accomplished here.

In conclusion, it should be stressed that the studies reported here are not offered as a way of
settling long-standing epistemological debates and that, in demonstrating a difference between
representational and constructivistic theories of mind, I make no claims whatsoever about that
wider discussion. While the data generated by this sequence of studies are meant to be all about
how young persons think about the knowing process, they are still not touted as being directly
relevant to the question of whether children or others actually do go about the business of having a
mental life by employing mental representations (built constructively or otherwise), or, meet the
even more demanding standard of explicitly realizing that they do so. In addition, the results
should not be taken to support an updated “two-miracle” claim that holds that children acquire their
last theory of mind at the age of 7 rather than 4. Understanding that different persons may
differently interpret droodles is still a very long way from a more mature and thorough-going
interpretive view and from the strife that attends our adult encounters with the knowledge that
knowledge is a personal construction.


APPENDICES

Appendix I: Posting sent to USENET newsgroups on the Internet in order to recruit subjects for Study 1. ..............................................................p.122

Appendix II: Sample item from the Internet questionnaire used in Study 1.................p.123

Appendix III: Stimulus Pictures used in Studies 1–6..................................................p.125

Appendix IV: Art and Interpretation .................................................................p.133
Appendix I: Posting sent to USENET newsgroups on the Internet in order to recruit subjects for Study 1.

From: chrisl@wimsey.com (Chris Lalonde)
Newsgroups: sci.psychology.research, sci.psychology.theory,
             sci.psychology.misc, sci.psychology.consciousness,
             sci.psychology.announce
Subject: Help a poor graduate student

I need a sample of adults to view some visual stimuli I've been using to study children's understanding of "interpretation." This is part of my doctoral thesis. If you have access to the World Wide Web, please take a few minutes to visit the following site for more information about an on-line questionnaire:

<http://vanbc.wimsey.com/~chrisl/Help.html>

No hidden agenda or marketing survey tricks. I just need adults to view 7 little line drawings and to type a few words to describe each picture.

Thanks for helping,

-Chris

Chris Lalonde <chrisl@wimsey.com>
Dept. of Psychology, University of British Columbia
Appendix II: Sample item from the Internet questionnaire used in Study 1.

A portion of Version A of the on-line questionnaire appears below. The instructions for both Version A and B are identical. It should be noted that in Version B, the item shown on the following page would appear without the small graphic that shows the full picture and the text would read: “What will Ann think this is?”.

**Instructions:** This questionnaire is based on a procedure used with 5- to 8-year-old children. Before seeing the pictures below, the children are given the following instructions:

Here are two puppets, Raggedy Ann and Raggedy Andy. We’re going to pretend that they’re real people like you and me. Ann and Andy have never seen any of these pictures before [I show them one of the pictures]. We’re going to show them this picture like this [I place a cover over the picture leaving only a small portion still visible through a window in the cover]. Then I bring out the puppets, one at a time, and ask the child what each puppet will think.

Below are each of the 7 pictures the children would see when the cover is closed and the small viewing window is opened. Please look at each of the 7 pictures below and then describe in a few words what “Ann” and “Andy” might think of each picture.

**Note:** If you have not read the CONSENT PAGE, please do so now.
PICTURE 1:

1A. What will **Ann** think this is? (righthand picture above)

1B. What will **Andy** think this is? (righthand picture above)

First Name: ___________________________ Age: ________

E-Mail Address: ___________________________

Sex: ☐ Male ☐ Female

Comments:

Note: By submitting your responses using the button below, it will be assumed that you have given your consent to participate in this study.
Appendix III: Stimulus Pictures used in Studies 1–6.

Label: Witch

Adapted from the droodle by Price that carried the caption:

"A ship arriving too late to save a drowning witch."
Label: Tower
Label: Cow

From a cartoon by Sandra Boyton
Label: Shark

From a *Farside* cartoon by Gary Larson.
Label: Pooh

Taken from an A.A. Milne illustration of "Winnie the Pooh"
Label: Giraffe
Label: Elephant

Adapted from the droodle by Price that carried the caption:

"An elephant smelling a grapefruit"
Label: Snake
Appendix IV: Art and Interpretation

Wrighter's Block

by Stuart McKenzie

I DON'T UNDERSTAND THIS PAINTING.

DON'T YOU SEE... IT'S THE ARTIST'S WAY OF SAYING THAT WE PROJECT OUR OWN VISION OF REALITY ONTO THE CANVAS OF LIFE.

...THAT WE ALL PERCEIVE REALITY IN A DIFFERENT WAY.

YOU MEAN IT'S NOT MICHAEL JACKSON IN A SNOWSTORM?

ARTISTS OFTEN USE SYMBOLISM IN THEIR WORK TO CONVEY A DEEPER MESSAGE.

FOR EXAMPLE, WHAT DOES THIS PIECE SAY TO US?

IT SAYS FIRE EXTINGUISHER PERHAPS IT'S A COMMENT ON OZONE DEPLETION

I THINK MY FAVOURITE PIECES IN THE ART SHOW WERE THE ABSTRACTS BY PICASSO.

I REALLY RELATED TO THE RAW, EMOTIONAL BRUSH STROKES OF VAN GOGH.

I WAS IMPRESSED BY THE METICULOUS PRECISION OF LEONARDO DA VINCI.

I STILL LIKE THE WASCALLY WABBIT ON MY T-SHIRT.